Journal of Classification 31:242-269 (2014)
DOI: 10.1007/s00357-014-9154-y

Using Generalized Procrustes Analysis for Multiple Imputation in Principal Component Analysis

Joost R. van Ginkel

Leiden University, The Netherlands

Pieter M. Kroonenberg

Leiden University, The Netherlands

Abstract: Multiple imputation is one of the most highly recommended procedures for dealing with missing data. However, to date little attention has been paid to methods for combining the results from principal component analyses applied to a multiply imputed data set. In this paper we propose Generalized Procrustes analysis for this purpose, of which its centroid solution can be used as a final estimate for the component loadings. Convex hulls based on the loadings of the imputed data sets can be used to represent the uncertainty due to the missing data. In two simulation studies, the performance of Generalized Procrustes approach is evaluated and compared with other methods. More specifically it is studied how these methods behave when order changes of components and sign reversals of component loadings occur, such as in case of near-equal eigenvalues, or data having almost as many counterindicative items as indicative items. The simulations show that other proposed methods either may run into serious problems or are not able to adequately assess the accuracy due to the presence of missing data. However, when the above situations do not occur, all methods will provide adequate estimates for the PCA loadings.

Keywords: Convex hulls; Missing data; Multiple imputation; Principal component analysis; Generalized Procrustes analysis; Questionnaires.

Corresponding Author's Address: J.R. van Ginkel, Child and Family Studies, Leiden University, Wassenaarseweg 52, 2333 AK Leiden, The Netherlands, tel: +31-(0)71-527 3620, email: jginkel@fsw.leidenuniv.nl.

Published online: 9 July 2014

1. Introduction

Multiple imputation (Rubin 1987) is a general procedure for dealing with missing data, which can be used for a wide variety of statistical techniques. The technique consists of four steps: (1) Missing values are estimated M times according to a specific statistical model; (2) These estimates are substituted into the data set, resulting in M plausible complete versions of the incomplete data set; (3) Standard statistical procedures are applied to these M data sets; (4) The results are pooled to obtain estimates for the parameters and their variability.

Relatively little attention has been given in the literature to multiple imputation in principal component analysis (PCA; for an extensive discussion of PCA, see Jolliffe 2002). The major difficulty for multiple imputation in PCA is the pooling of the results from the analyses of the imputed data sets to obtain a single estimate of the loading matrix. While averaging is an appropriate pooling procedure in many other statistical techniques, this is not recommended for PCA due to the lack alignment of the components from separate solutions. It is necessary to optimally align component spaces first before they can be combined. In this paper it is shown that Generalized Procrustes Analysis (Gower 1971; 1975; Ten Berge 1977) is ideally suited for this purpose. Several alternatives, like averaging the correlation matrices of the imputed data sets and averaging (varimax-rotated) component loading matrices are discussed and studied as well.

The exposition will be set in the context of research using psychological tests and questionnaires given their ubiquitousness in the social and behavioral sciences. Generally tests and questionnaires are highly structured and multidimensional, which allows for a clear evaluation of the combination techniques.

2. Missingness Mechanisms and Alternative Methods for Missing Data in PCA

To better understand the appropriateness of multiple imputation in the context of PCA, it is necessary to discuss the missingness mechanisms that are described in the literature, along with other methods for handling missing data. The literature on missing data (e.g. Little and Rubin 2002; Schafer 1997; Rubin 1976) distinguishes three missingness mechanisms: Missing Completely At Random (MCAR), Missing At Random (MAR), and Not Missing At Random (NMAR).

When missing data are MCAR missing values are a simple random sample of all data points. When missing data are MAR they depend on observed variables. This can be a background variable such as gender (e.g., men leave more questions unanswered than women) or an item in the questionnaire (e.g., higher scores on this item concur with higher probabilities that a respondent leaves other questions unanswered). Finally, when data are NMAR the missing data depend on information that is not observed. This can be an unobserved variable (for example, missingness depends on gender, gender being unobserved), or the value of the missing score itself (e.g., high scores having a higher probability of being missing than lower scores). For a more technical description of the three missingness mechanisms, we refer to the above-mentioned literature.

Traditional methods for handling missing data in PCA such as listwise and pairwise deletion are only guaranteed to give unbiased results when the missingness mechanism is MCAR. Since this is a very strong assumption, several other methods (along with multiple imputation) have been put forward that will also lead to valid results under MAR mechanisms. The major proposals in the present context are Weighted least squares fitting (Kiers 1997; Grung and Manne 1997), Regularized PCA (Josse, Pagès, and Husson 2011; 2009), the EM algorithm (Bernaards and Sijtsma 1999; 2000; Schafer 1997, pp. 163–181), Maximum likelihood principal component analysis (Wentzell, Andrews, Hamilton, Faber, and Kowalski 1997), the Missing-data passive method (Benzécri 1973; Meulman 1982; Takane and Oshima-Takane 2003), and Symbolic PCA (Zuccolotto 2008). One of the drawbacks of these procedures compared to multiple imputation is that they will only lead to unbiased results for specific types of MAR. More specifically, these methods cannot handle missing data that are related to background variables outside the model (such as gender) because for handling the missing data they only use the variables included in the PCA. Thus, even though the data are MAR, for the above-mentioned methods this assumption is violated, and valid results can no longer be guaranteed.

Multiple imputation on the other hand, can use all available information in the imputation model, including background variables, so that the relations between the missing data and the background variables are also accounted for in the imputed values, and in the subsequent PCA. Therefore, multiple imputation will lead to unbiased results for any type of MAR, provided the imputation model sufficiently resembles the model underlying the data. The multivariate normal model (Schafer 1997, Chap. 5) has shown to be adequate in many cases (Graham and Schafer 1999; Bernaards, Belin, and Schafer 2007).

Finally, under NMAR no method is guaranteed to give unbiased results. However, Schafer (1997, pp. 26–28) argued and showed that the more variables are included in the imputation model, the more likely it is that the MAR assumption holds. Thus, since multiple imputation can use a larger number of variables for handling the missing data than the above-mentioned methods, it will also be more robust to departures from MCAR and MAR.

3. Multiple Imputation and PCA

In the previous section it was argued that multiple imputation resolves the problem of dependence of the missingness on observed covariates outside the PCA model. However, using multiple imputation in PCA creates another problem, i.e. how the results from several imputed data sets can be combined into one overall PCA solution. Rubin (1987, p. 2) proposed several procedures for pooling standard errors, parameter estimates, and significance tests resulting from several imputed data sets. Few such procedures have been proposed for pooling the results from PCA apart from averaging loading matrices (see, for example, Ho, Silva, and Hogg 2001; Masi, Aldag, and Chatterton 2006) and averaging the correlation matrices of the imputed data sets prior to performing a PCA on the average (Van Ginkel, Van der Ark, Sijtsma, and Vermunt 2007; Van Ginkel 2010).

3.1 Averaging Component Loadings

With the unavailability of an adequate pooling technique for PCA in multiple imputation, the first thing a researcher may think of is to average loading matrices. With only M separate PCA solutions from M imputed data sets at hand, this is the only option that can be performed manually. Because generally *Varimax-rotated* solutions are interpreted rather than the unrotated solutions, it seems most convenient to average the *Varimax-rotated* solutions. The resulting method will henceforth be referred to as the Mean Varimax method (MVM).

MVM has three disadvantages. Firstly, averaging gives incorrect results when the order of the components is not the same for all imputed data sets. For example, in an incomplete data set with 10 items with two components with near equal variance, it may happen that in one imputed data set items 1-5 load highest on the first component, and items 6-10 load highest on the second component, while for another imputed data set this may be reversed. Secondly, in one or more imputed data sets the signs of the loadings may be reversed compared to those of other imputed data sets. This may happen when a questionnaire contains about equally many indicative as counterindicative items. A third disadvantage is that the Varimax rotated solutions of the M imputed data sets may have been rotated according to the Varimax criterion, but are not optimally rotated towards each other. As a result, the average solution is computed across solutions that have more variation among each other than necessary (for a discussion of these problems, see, e.g., Chatterjee 1984; Markus 1994; Milan and Whittaker 1995; and Linting, Meulman, Groenen, and Van der Kooij 2007). A better pooling procedure would be to put the unrotated components of all imputed data sets in the same order, switch signs for some sets of the unrotated loadings, optimally rotate the solutions towards each other and then compute an average component solution. Finally, for interpretational purposes this average solution may then be rotated according to the Varimax criterion.

3.2 PCA of the Mean Correlation Matrix

Van Ginkel et al. (2007) and Van Ginkel (2010) investigated the influence of multiple imputation on the results of PCA by first computing the average correlation matrix based on several imputed data sets, followed by a PCA with a Varimax rotation. This approach will henceforth be referred to as mean correlation matrix (MCM). This method resolves the problem of order change and sign reversal of loadings that MVM method has, since pooling takes place at the level of the correlation matrix rather than at the level of the Varimax-rotated component loadings. However, when loadings are computed for an average correlation matrix, all information about the variability among imputed data sets is lost. As a result it becomes impossible to get an impression of the uncertainty of the PCA solution due to the missing data.

3.3 Generalized Procrustes Analysis for Pooling Component Loadings

Generalized Procrustes analysis (GPA; Gower 1971; 1975; Ten Berge 1977) solves all three aforementioned problems (order change, sign reversal and rotational freedom of PCA solutions with respect to each other) at the same time by producing an optimal average solution, in particular the centroid of the PCA solutions from the imputed data sets as well as an assessment of the amount of uncertainty due to the missing data. The same idea for multiple imputation in PCA has already been discussed by Kroonenberg (2008, pp. 152–153) and by Van Ginkel and Kroonenberg (2009), inspired by a comment of D'Aubigny (2004). See also Josse, Pagès, and Husson (2011) who used a similar idea: they applied a PCA to the incomplete data using Weighted least squares (Kiers 1997). Next they multiply imputed the data under this PCA model, and used a Procrustes rotation (see details below) to rotate each solution optimally to the WLS solution.

GPA in the context of multiple imputation works as follows. Suppose $\mathbf{A}_m$ ($m = 1,\ldots, M$) is a $J \times K$ matrix containing the principal component loadings of the m-th imputed data set, where J is the number of variables and K the number of components. In GPA we want to find the $K \times K$ orthogonal transformation matrix $\mathbf{T}_m$ for each of the M imputed data sets such that the sum of the squared distances between the transformed loading matrices is minimized.

$$f(\mathbf{T}_1,...,\mathbf{T}_M) = \sum_{i<j} tr(\mathbf{A}_i\mathbf{T}_i - \mathbf{A}_j\mathbf{T}_j)'(\mathbf{A}_i\mathbf{T}_i - \mathbf{A}_j\mathbf{T}_j)\,. \tag{1}$$

To find the transformation matrices $\mathbf{T}_1,\ldots,\mathbf{T}_M$ we must first consider the special case when there are only two matrices **A** and **B** and **A** needs to be optimally fitted to **B**. This is known as the classical orthogonal Procrustes problem (see Green 1952; Gower 1971). This problem requires only one transformation matrix **T** which may be found as follows: suppose $\mathbf{QLV'}$ is a singular value decomposition of matrix $\mathbf{A'B}$. Then the transformation matrix **T** can be found by means of

$$\mathbf{T} = \mathbf{QV'} \tag{2}$$

By multiplying **A** with **T**, **A** is optimally fitted to **B**.

For more than two matrices, Ten Berge's (1977, p. 272) procedure may be used. Define t as the iteration number. Starting at t = 1, the algorithm goes as follows:

Step 0: Set $\mathbf{T}_m = \mathbf{I}$ for $m = 2,\ldots, M$.

Step 1: Fit $\mathbf{A}_1$ optimally to $\mathbf{B} = \sum_{m=2}^{M}\mathbf{A}_m\mathbf{T}_m$ using transformation matrix **T** as computed in Equation 2, yielding $\mathbf{A}_1\mathbf{T}_1^{(t)}$.

Step 2: Fit $\mathbf{A}_2$ optimally to $\mathbf{B} = \mathbf{A}_1\mathbf{T}_1^{(t)} + \sum_{m=3}^{M}\mathbf{A}_m\mathbf{T}_m$ yielding $\mathbf{A}_2\mathbf{T}_2^{(t)}$.

…

Step M: Fit $\mathbf{A}_M$ optimally to $\mathbf{B} = \sum_{m=1}^{M-1}\mathbf{A}_m\mathbf{T}_m^{(t)}$, yielding $\mathbf{A}_M\mathbf{T}_M^{(t)}$.

Step M+1: Fit $\mathbf{A}_1\mathbf{T}_1^{(t)}$ optimally to $\mathbf{B} = \sum_{m=2}^{M}\mathbf{A}_m\mathbf{T}_m^{(t)}$, yielding $\mathbf{A}_1\mathbf{T}_1^{(t+1)}$.

Steps 2-M are repeated, increasing t with 1 at each iteration, until convergence.

Once convergence has been achieved, the centroid configuration (the mean of all transformed solutions) is used as the final solution for the M imputed data sets. For interpretational purposes, the centroid solution may be rotated either with an orthogonal or an oblique transformation. A more detailed discussion of GPA may be found in Commandeur (1991), Ten Berge (1977), and Gower (1975).

3.4 Uncertainty About the PCA Solutions

One important advantage of multiple imputation over single imputation is that it takes statistical stability into account. This usually means that standard errors and confidence intervals of parameter estimates are adjusted for the extra uncertainty caused by the missing data (Rubin 1987). Although it is possible to construct confidence intervals for component loadings (e.g., Girshick 1939; Anderson 1963; Archer and Jennrich 1973; Ogasawara 2000; 2002; Timmerman, Kiers, and Smilde 2007), standard PCA is mostly used for descriptive purposes, and statistical tests or confidence intervals are rarely used.

However, also for descriptive PCA it is possible to show the uncertainty about the sample loadings as a result of the missing data by constructing a loading plot for all imputed PCAs simultaneously. Each variable is represented by M points, one for each PCA. Per variable, for M loadings the convex hull is computed resulting in a polygon which contains all M solutions. The surface of such a convex hull gives an impression of the uncertainty due to the missing data of the estimated component loadings. It should be noted that the convex hulls only give relative information about the uncertainty of the loadings, namely how much variability a loading has relative to the others. A loading with a large convex hull suffers much from the missing data so that consequently we have to be cautious with its interpretation. However, this relative information fits well into the explorative nature of PCA, in which usually no statistical inferences are drawn about parameter estimates. Furthermore, in other contexts than multiple imputation it is also common to use convex hulls in an exploratory manner (see, e.g., Rousseeuw, Ruts, and Tukey 1999). If one wants absolute information about the uncertainty of the loadings in the form of confidence intervals, one could turn to Convex hull peeling (Green 1981) or confidence ellipses (e.g., Josse, Pagès, and Husson 2011). Figure 1 shows an example of such a loading plot for a simulated data set from a simulation study that will be discussed later on.

The larger the surface of a convex hull, the more uncertain the exact the location of the loading becomes as a result of the missing data. Thus, when items have low centroid loadings but large convex hulls, the items are not necessarily of bad quality but the missing data may simply introduce too much uncertainty about the exact values of the loadings.

3.4.1 Theoretical Properties of the Convex Hulls

The question is to what extent the convex hulls will cover the values of the true sample loadings if the data had been complete. To answer this question it is important to introduce the concept of a *Bayesianly proper* imputation method (Schafer 1997, p. 105). Suppose $\mathbf{X}$ is an incomplete data matrix with a missing part $\mathbf{X}_{mis}$ and an observed part $\mathbf{X}_{obs}$, so that $\mathbf{X} = (\mathbf{X}_{mis}, \mathbf{X}_{obs})$, and that θ is a set of model parameters. An imputation method is said to be *Bayesianly proper* if its imputations are independent realizations of the posterior distribution of $\mathbf{X}_{mis}$ given $\mathbf{X}_{obs}$ and parameter

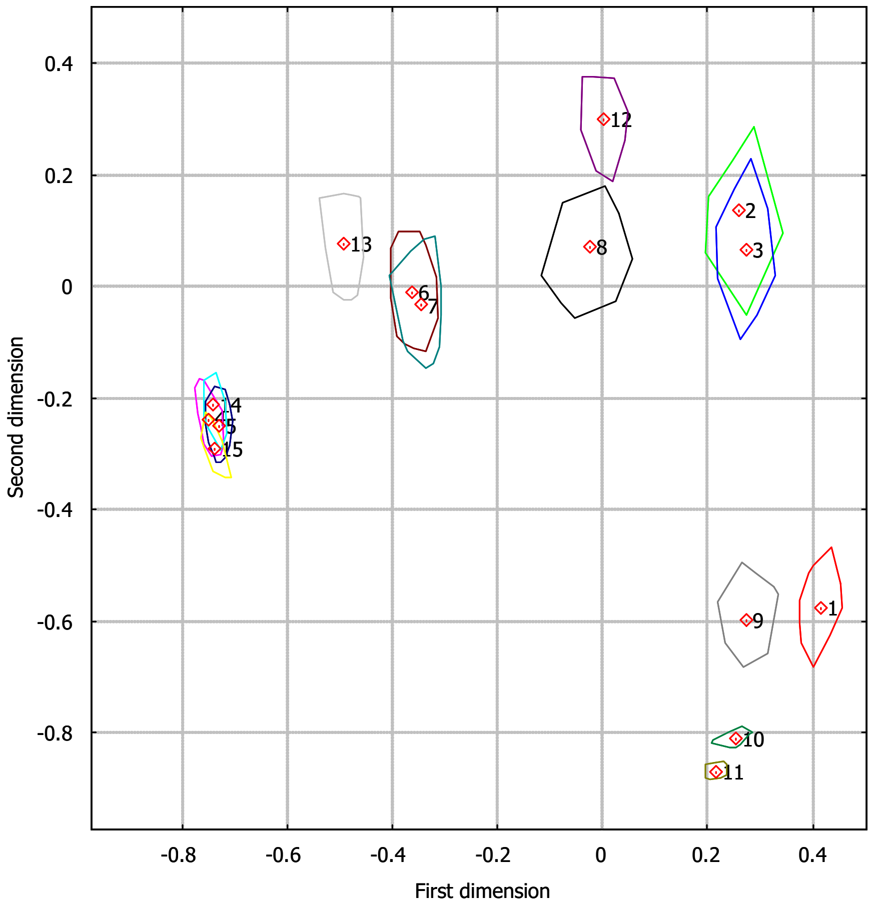

Figure 1. Loading plots With Convex Hulls of Component Solution of a Simulated Data Set, Using Generalized Procrustes Analysis.

set θ, and given that the missingness is MAR. This posterior distribution, denoted $P(\mathbf{X}_{mis} \mid \mathbf{X}_{obs})$, can then be written as:

$$P(\mathbf{X}_{mis} \mid \mathbf{X}_{obs}) = \int P(\mathbf{X}_{mis} \mid \mathbf{X}_{obs}, \theta) P(\theta \mid \mathbf{X}_{obs}) d\theta . \tag{3}$$

The above definition implies that θ has to apply for both the model that generated the data and the imputation model. In other words, the imputation model must equal the model that generated the data.

Another implication of the above definition is that the data have to be MAR. If the data are NMAR, then $\mathbf{X}_{obs}$ does not accurately represent the data and consequently, draws from the posterior $P(\theta \mid \mathbf{X}_{obs})$ will not be representative of $P(\theta \mid \mathbf{X})$. As a result, draws from the distribution of $P(\mathbf{X}_{mis} \mid \mathbf{X}_{obs})$ possibly covers only part of the total possible values for $\mathbf{X}_{mis}$. However, if all of the above conditions are met, it follows automatically that for an infinite number of imputations the true sample loading is always covered by the convex hull, since the true value that would have been obtained if the data had been complete, is just one of the possible draws from the distribution of $P(\mathbf{X}_{mis} \mid \mathbf{X}_{obs})$.

In practice the convex hulls will not (always) capture all the loadings that would have been obtained if the data had been complete. There are two possible reasons for this. The first reason is that the imputation model deviates from the model that generated the data. It is therefore important to use an imputation model that accurately describes the data. The second reason is that the number of imputations is finite. The question is how many imputations are needed for the convex hulls to capture a fair amount of the true sample loadings, which is also one of the research questions of this study.

4. Comparing Procedures for Finding a Pooled Component Solution

On theoretical grounds, GPA is superior for handling principal component analysis solutions from multiply imputed data sets because it solves all three problems mentioned earlier (order change, sign reversal and rotational freedom) simultaneously. However, it is an open question how strong such an advantage is in well-behaved practical applications compared MVM.

To answer this question, two simulation studies were carried out. In the first simulation study, incomplete data sets were drawn from artificial populations and were multiply imputed. Component solutions resulting from GPA, MVM, and MCM were compared with the component structure of the corresponding artificial populations.

In the second simulation study the focus was the uncertainty about the PCA loadings represented by convex hulls. More specifically, the purpose was first of all to show how much larger the surfaces of the convex hulls become when convex hulls are not corrected for sign reversals, order changes, and rotational freedom. The second purpose was to see whether the convex hulls capture a fair amount of the loadings of the original sample data. Here, GPA was compared with MVM. MCM was not considered because it throws away information about the uncertainty due to the missing data.

In the next section, the simulation procedures and the designs of the simulation studies are discussed in detail, followed by their results. Finally, conclusions are drawn about the performance of MVM, MCM, and GPA.

5. Simulation Study 1

5.1 Method

For the first simulation study, three populations were constructed. Instead of using standard theoretical distributions, the populations were based on three real data sets, denoted the *population data sets*. This approach was used in order to more closely mirror situations that may occur in reality (for details see below). The data sets to be used in the simulations, denoted as the *original sample data sets*, were randomly drawn samples from the three population data sets. Subsequently, different percentages of missingness were created in these data sets, resulting in what we will call *incomplete sample data sets*. Next, the missing scores were estimated several times using multiple imputation, and the resulting completed data sets are referred to as *imputed sample data sets*. For the imputed sample data sets component solutions were combined either by using GPA, MVM, or MCM. For each final component solution, it was established via a Euclidean loss function and a measure for bias how closely the combined component solution resembled the component structure based on a PCA on the population data set.

5.1.1 Simulating the Population Data Sets

Two of the population data sets were constructed using two real data sets from the NICHD (1996) data. The third population data set was constructed using a data set from a study about posttraumatic stress disorder (PTSD) in children (Alisic, Van der Schoot, Van Ginkel, and Kleber 2008). This data set has also been used in another multiple imputation study (Van Ginkel 2010).

The first NICHD data set consisted of the scores of 1016 mothers on the Child-Parent Relationship Scale (Pianta 1992). The questionnaire consists of 15 items all of which had five-point response scales. All items were constructed to measure mothers' perceptions about the relationship with their children. Each item belonged to one of two subscales: Mother Conflict With Child (Items 2, 8, 10, 11, 12, 13, and 14) and Mother Closeness With Child (Items 1, 3, 4, 5, 6, 7, 9, and 15).

The total score on the 21-item Maternal Separation Anxiety Scale (Hock 1984) was used as a covariate for simulating the missing data. The probability that item scores on items of the Child-Parent Relationship Scale were missing increased linearly with the covariate score. For the highest covariate score the probability of item scores being missing was three times as high as for the lowest covariate score. This procedure corresponds to a missing-at-random mechanism (Rubin 1976). It is important to note that this covariate was part of the imputation model but was not included in the subsequent PCAs. Because methods such as the Missing-data passive approach (Benzécri 1973; Meulman 1982; Takane and Oshima-Takane 2003) and Weighted least squares fitting (Kiers 1997; Grung and Manne 1998) only use variables included in the PCA for handling the missing data their results may be biased as already mentioned in the introduction.

The second data set consisted of the responses of 1278 mothers to the Self Scale. The items on this scale are taken from the NEO Personality Inventory (Costa and McCrae 1985). The inventory consists of 36 items scored on five-point scales, and it has three subscales: Neuroticism (items 1, 4, 7, 10, 13, 16, 19, 22, 25, 28, 31, and 34), Extraversion (items 2, 5, 8, 11, 14, 17, 20, 23, 26, 29, 32, and 35), and Agreeableness (items 3, 6, 9, 12, 15, 18, 21, 24, 27, 30, 33, and 36).

This time the covariate score was the total scale score on the My Feelings scale (CES-D; Radloff 1977) which consists of 20 items scored on five-point scales, measuring maternal depression. This covariate score was used to simulate the same missing-at-random mechanism as for the first data set.

Finally, the third data set consisted of 1515 complete response patterns of children to the KIDSCREEN-27 (Ravens-Sieberer et al. 2007). This is a 27-item questionnaire that covers five quality-of-life dimensions. The complete data set is described in Alisic et al. (2008). Here, gender was the covariate on which the missing data depended. For girls the probability of being missing was three times as high as for boys.

All empirical data sets were too small to draw a large number of samples from with replacement of reasonable size and with a sufficiently small overlap. Therefore the three data sets were artificially enlarged to create three population data sets, as explained in the Appendix. The resulting populations are denoted Population 1 (Child-Parent Relationship Scale), Population 2 (Self Scale), and Population 3 (KIDSCREEN-27).

5.1.2 Simulating the Original Sample Data Sets

From each of the above-mentioned populations, 200 replicated data sets were drawn, 100 of which were of size $N = 200$ and 100 of which were of size $N = 500$. Different sample sizes were studied to see whether the performance of the pooling methods would be influenced by sample size.

5.1.3 Simulating the Incomplete Sample Data Sets

In each of the simulated original sample data sets, 5%, 10% and 15% of the item scores were removed. Because missingness mechanism was not the topic of this paper, it was decided to use only one mechanism for simulating the missingness. The MAR mechanism described earlier was chosen for this purpose because MCAR is an ideal situation which is not likely to occur in practice, while NMAR may introduce additional bias in the analyses, and may obscure the comparisons.

5.1.4 Imputing the Incomplete Sample Data Sets

The resulting incomplete sample data sets were imputed using multiple imputation under the multivariate normal model (Schafer 1997, Chap. 5). Dependence of the missingness on the covariate was taken into account by including the covariate in the imputation model.

One reason for using this imputation method is that it is widely available in many statistical software packages such as NORM (Schafer 1998), the missing-data library of S-plus 8 for Windows (2007), the mi package in R (Su, Gelman, Hill, and Yajima 2011), and SAS 8.1 in the procedure PROC MI (Yuan 2011). The statistical package SPSS (SPSS Inc. 2011), can also perform multiple imputation under this model, but uses a different computational technique (MICE; Van Buuren, Brand, Groothuis-Oudhoorn, and Rubin 2006).

A second reason for using the multivariate normal model is that it is most appropriate for statistical techniques that assume continuous data like PCA. In this paper rating scales are considered discretized continuous numeric variables, which is in line with the prerequisites for standard principal component analysis. There has been considerable discussion on the appropriateness of treating rating scales as such (For a summary of the arguments see, e.g. Knapp 1990; Doering and Raymond 1979; Acock, and Martin 1974), but it is common practice to do this. Besides, there is some evidence that discrete rating scales can safely be treated as interval scales (Nandakumar, Yu, Li, and Stout 1998; Bollen and Barb 1981; also, see Labovitz 1967; Baker, Hardyck, and Petrinovich 1966).

Note that the imputation model makes an even stronger assumption, namely that sets of rating scales can be conceived of as multivariate normally distributed. However, multiple imputation under the multivariate normal model was found to be robust against departures from the multivariate normal model (Graham and Schafer 1999; Bernaards, Belin, and Schafer 2007).

For this study, data were imputed five times. Early literature states that this number of imputations is sufficient for most purposes (see, Schafer 1997, pp. 106–107; pp. 197–199). However, more recent literature (for an overview, see Van Buuren 2012, pp. 49–51) states that more imputations may be needed. For our purposes a larger number was not necessary because additional uncertainty as a result of a small number of imputations cancels itself out across replicated data sets.

Finally, PCAs were carried out for the resulting imputed data sets and combined using GPA and MVM; for MCM the correlation matrices were pooled (see above).

5.1.5 Dependent Variables

For each pooled PCA solution the following fit index served as a quality measure of the solution. Suppose a_{jk} is the population component loading of item j on component k, and $\hat{a}_{jk,v}$ is the corresponding centroid/mean loading of M imputed versions of incomplete sample data set v. The root mean squared bias $RMSB_v$ for the v-th incomplete data set is defined as

$$RMSB_v = \sqrt{\sum_{j=1}^{J}\sum_{k=1}^{K}(\hat{a}_{jk,v} - a_{jk})^2 / JK} .$$

The value of $RMSB_v$ can be interpreted as the root of the squared distance of the loadings of either the original sample data set v (when computed for the original sample data sets) or the multiply imputed sample data set v (when computed for the imputed sample data sets) to the population loadings in the multidimensional space, averaged over J items and the K extracted components. The smaller $RMSB_v$, the better the component solution of the original or imputed sample data fits the component structure of the population (loadings of the imputed data sets fit the loadings of the population data set).

Besides this fit measure, the non-squared distance of $\hat{a}_{jk,v}$ from a_{jk} was also studied because this gives an indication of whether loadings are systematically under- or overestimated. That is, for each original sample data set v and multiply imputed sample data set v the mean bias, denoted MB, across items and components was computed as:

$$MB_v = \frac{1}{JK}\sum_{j=1}^{J}\sum_{k=1}^{K}(\hat{a}_{jk,v} - a_{jk}) .$$

The MB_v averaged across replications serves as an estimate for bias across items and components.

Prior to computing $RMSB_v$ and MB_v, component solutions of both imputed and original sample data sets were optimally rotated towards the population solution using an orthogonal Procrustes rotation (Green 1952; Gower 1971) to correct for rotational freedom.

5.2 Results

For each of the populations, two 2 (Sample size) × 3 (Percentage of missingness) × 2 (Pooling technique) ANOVAs were carried out with the $RMSB_v$ and MB_v as the dependent variables. We used two ANOVAs instead of one MANOVA because even though $RMSB_v$ and MB_v are both quality measures, they are quite different concepts with different properties. Following Cohen's (1988) guidelines for effect sizes, small (.01 ≤ total η^2 < .06), medium (0.6 ≤ total η^2 < .14), and large effects (total η^2 ≥ .14) of the ANOVAs are reported. Table 1 shows the ANOVA results for the significant effects with a discernable size.

The table shows that for all populations, Sample size (N) and Percentage of missingness met Cohen's criteria for effect sizes for $RMSB_v$; for MB_v only Sample Size had a small effect in Population 2. Pooling technique had a discernable effect size for Populations 2 and 3. Finally, the interaction of Sample size and Pooling method was small for Population 3.

In general, as the percentage of missingness increased, both $RMSB_v$ and MB_v increased. Furthermore, for sample sizes of N = 500 $RMSB_v$ was smaller than for sample sizes of N = 200. This makes sense because a larger sample size increases the stability of the PCA solution, resulting in smaller deviations from the population solutions. For Populations 2 and 3 there were discernable effects of Pooling technique on $RMSB_v$. GPA produced on average the smallest value of $RMSB_v$, followed by MCM, and MVM. When inspecting the data, it turned out that this large $RMSB_v$ for MVM was indeed due to sign reversals and order changes of components. Finally, the interaction of Sample size and Pooling technique can be described as follows: as the sample size increased, the value of $RMSB_v$ decreased for all methods, only for MVM it decreased less rapidly than for GPA and MCM.

Table 2 shows the means (M) and standard deviations (SD) of $RMSB_v$ for all combinations of sample size, missingness mechanism, for pooling technique and for all populations. The results of the original sample data (without missing values) are shown in the first row (0% missingness), for comparison with the results of the imputed sample data. The table shows that $RMSB_v$ has similar magnitude for methods GPA, and MCM and that MVM generally has larger values of $RMSB_v$. The results for the mean bias MB_v are shown in Table 3. Like $RMSB_v$, results are very similar for MCM and GPA, and the results of MVM deviate from those of MCM and GPA.

Furthermore it can be seen from Tables 2 and 3 that on average neither of the methods GPA and MCM clearly performed best. With respect to $RMSB_v$ (Table 2) GPA performed best, whereas for MB_v (Table 3) MCM performed best.

Table 1. ANOVA results for mean bias and root mean squared bias of effects with a discernable size.

Number of components	Dependent variable	Independent variable	F	$df1$	$df2$	η^2
2	RMSB	N	328.87	1	198	.59***
		Percentage missingness	68.41	2	396	.01*
	MB	N	6.17	1	198	.03*
3	RMSB	N	577.37	1	198	.62***
		Percentage missingness	9.74	2	396	.02*
		Pooling technique	54.52	2	396	.02*
5	RMSB	N	579.41	1	198	.60***
		Percentage missingness	175.65	2	396	.03*
		Pooling technique	322.91	2	396	.06*
		$N \times$ Pooling technique	106.73	2	396	.02*

All p-values were less than .001, except for the effect of Sample size on MB for 2 components ($p = .014$).

Table 2. Results for root mean squared bias, shown for all percentages of missingness, pooling techniques, and populations. Totals are aggregated across pooling technique. Entries must be multiplied by 10^{-3}.

N	% missing	Pooling technique	Population1		Population 2		Population 3	
			M	SD	M	SD	M	SD
200	0%	-	264.61	23.89	274.15	20.57	310.94	22.95
	5%	GPA	270.93	25.07	279.14	22.55	317.91	23.07
		MCM	271.03	25.10	279.34	22.65	319.21	24.24
		MVM	270.89	24.99	291.31	36.67	328.92	20.92
	10%	GPA	276.94	28.33	286.02	23.76	324.16	23.77
		MCM	277.19	28.39	286.47	24.11	325.54	24.59
		MVM	276.80	28.22	298.29	34.91	334.44	18.40
	15%	GPA	283.44	27.80	293.36	25.31	333.92	23.23
		MCM	284.78	27.42	294.12	25.57	336.16	24.61
		MVM	283.55	27.18	309.77	37.43	340.73	19.10
500	0%	-	209.32	19.51	213.04	13.45	242.60	15.95
	5%	GPA	214.11	20.07	216.90	12.80	248.81	18.94
		MCM	214.12	20.06	216.92	12.79	248.94	18.90
		MVM	214.11	20.07	221.93	25.51	274.08	25.07
	10%	GPA	218.98	20.42	221.67	12.96	256.01	18.59
		MCM	219.14	20.44	221.70	12.94	255.91	18.59
		MVM	218.97	20.42	230.39	29.54	290.82	23.45
	15%	GPA	223.30	21.97	226.61	12.61	265.13	21.73
		MCM	223.36	22.01	226.72	12.57	265.40	21.91
		MVM	223.27	21.98	238.80	22.32	300.01	22.88

Table 3. Results for mean bias, shown for all percentages of missingness, pooling techniques, and populations. Totals are aggregated across pooling technique. Entries must be multiplied by 10^{-3}.

N	% missing	Pooling technique	Population1		Population 2		Population 3	
			M	*SD*	*M*	*SD*	*M*	*SD*
200	0%	-	-4.11	16.90	-2.85	15.03	-0.33	8.28
	5%	GPA	-6.22	18.30	-3.65	14.97	-0.71	8.27
		MCM	-5.77	18.09	-3.53	14.96	-0.72	8.23
		MVM	-6.25	18.33	-4.56	15.20	-1.11	8.23
	10%	GPA	-7.73	19.01	-4.85	18.21	-1.13	8.34
		MCM	-7.08	18.76	-4.51	18.25	-1.19	8.33
		MVM	-7.87	19.31	-6.43	18.55	-1.54	8.23
	15%	GPA	-9.64	20.05	-6.25	18.61	-1.73	8.63
		MCM	-8.72	19.86	-6.09	20.04	-1.75	8.33
		MVM	-9.76	20.13	-8.02	18.85	-2.17	8.34
500	0%	-	-1.70	9.15	-1.46	8.94	-0.69	4.80
	5%	GPA	-2.19	9.60	-1.70	9.39	-0.84	4.85
		MCM	-2.10	9.58	-1.66	9.39	-0.84	4.85
		MVM	-2.20	9.60	-1.81	9.30	-0.51	5.32
	10%	GPA	-2.75	9.97	-1.91	9.55	-1.02	4.89
		MCM	-2.55	9.90	-1.81	9.54	-1.02	4.88
		MVM	-2.76	9.97	-2.21	9.44	-8.44	5.44
	15%	GPA	-2.64	10.10	-2.42	9.75	-1.43	4.98
		MCM	-2.31	9.99	-2.23	9.71	-1.44	4.97
		MVM	-2.66	10.10	-2.88	9.68	-1.64	5.99

One final noticeable result is that the mean bias for no missing data (Table 3, first row) is not equal to 0. However, it turned out that this could be due to sampling variability (one-sample *t*-tests with Bonferroni correction).

6. Simulation Study 2

6.1 Method

In this simulation study it was studied how often the convex hulls of the multiply imputed data captured the loadings of the corresponding original sample data in the two-dimensional space.

For this simulation study one of the populations from Study 1 was used, namely the Child-Parent-Relationship-Scale data (two dimensions). Using N = 200, the same original sample data sets were drawn from this population as in Study 1, and a fixed percentage of missing data was simulated, namely 15%.

6.1.1 Independent Variables

Two independent variables were studied, namely Pooling technique and # Imputations. Pooling methods were GPA and MVM. The # Imputations was set to 5, 20, and 100. Different numbers of imputations were used because as would turn out later, 5 imputations were not enough to capture a fair amount of the loadings of the original sample data.

6.1.2 Dependent Variables

Two dependent variables were studied. The first dependent variable was the amount of loadings of the original sample data that was captured by the corresponding convex hulls of the imputed data, as a percentage of the total number of items, denoted *percentage of captured loadings* (*PCL*). The second dependent variable was the average surface of the convex hulls of the component solution, denoted *average convex hull surface* (*ACHS*). The surface of one convex hull can be calculated as the sum of the triangles forming the convex hull. One way to calculate the surface of the triangles from their coordinates is by using Heron's rule dating back to before 200 BC. Suppose a triangle has three sides, namely side a, b, and c, and define $s = (a+b+c)/2$. The surface of one triangle is computed as $\sqrt{[s(s-a)(s-b)(s-c)]}$, and the total surface of the convex hull is the sum of all triangles (see Weisstein 2012).

6.2 Results

Two 2 (Pooling technique) × 3 (# Imputations) ANOVAs were carried out with *PCL* and *ACHS* as the dependent variables. The results of these ANOVAs can be found in Table 4. Means and standard deviations of *PCL* and *ACHS* can be found in Table 5.

Table 5 (first row) shows that for 5 imputations the percentage of loadings captured by the convex hulls (*PCL*) is only about 20% for both GPA and MVM. It was concluded that for 5 imputations the convex hulls are not very informative and thus it was decided to run the simulations with 20 imputations and 100 imputations as well (Table 5, second and third row). For 100 imputations the convex hulls capture a fair amount of the loadings, namely about 80%. For a test of 15 items (like in these simulations) this means that on average 13 out of 15 convex hulls include the corresponding loadings of the original sample data set.

The interaction of # Imputations × Pooling technique was not significant for the percentage of captured loadings but it was significant for the average convex hull surface, although it did not meet Cohen's criteria for effect sizes (Table 4). When looking at Table 5, this interaction may be interpreted as follows: as the number of imputations increases, the

Table 4. ANOVA for percentage of loadings captured by the convex hulls, average surface of the convex hulls.

Dependent variable	Independent variable	F	$df1$	p	η^2
PCL	Pooling technique	22.73	1	<.001	<.01
	# Imputations	1323.80	2	.01	.81***
	Pooling technique × # Imputations	0.35	2	.71	<.01
ACHS	Pooling technique	5.96	1	<.001	.01*
	# Imputations	67.06	2	<.001	.15***
	Pooling technique × # Imputations	4.23	2	.02	<.01

All error degrees of freedom are equal to $df2 = 198$.

Table 5. Results for PCL and ACHS, shown for all numbers of imputations, and poolting techniques. Totals are aggregated across number of imputations (rows), pooling technique (columns), or both (lower right corner in each panel). Entries for ACHS must be multiplied by 10^{-3}.

Dependent		GPA		MVM		Total	
variable	# Imputations	*M*	*SD*	*M*	*SD*	*M*	*SD*
PCL	5	.20	.09	.21	.10	.21	.1
	20	.59	.14	.61	.15	.60	.15
	100	.83	.12	.84	.12	.83	.12
	Total	.54	.29	.56	.29	.55	.29
ACHS	5	4	1	4	3	4	2
	20	13	5	18	20	16	15
	100	26	9	40	66	.33	47
	Total	14	11	21	42	18	31

average surface of the convex hulls also increases, but for GPA (first and second column) it increases less rapidly than for MVM (third and fourth column).

Table 4 (second row in both panels) shows that the largest effects were found for the number of imputations. As the number of imputations increased, both the percentages of captured loadings and the average convex hull surface increased (Table 5, last two columns). Pooling technique did not seem to be as influential. The effect of Pooling technique was small for the average convex hull surface (Table 4, first row of lower panel). As expected, GPA produced a smaller average surface than MVM did.

The effect of Pooling technique on the percentage of captured loadings did not meet Cohen's criteria for effect sizes, be it that MVM captured on average *more* loadings than GPA (Table 5, last row of upper panel). This is a counterintuitive result since it was expected that GPA would perform better than MVM. Thus, GPA produces convex hulls with smaller surfaces than MVM, but does not capture more loadings.

However, since the interaction of # Imputations × Pooling technique is not significant for *PCL* and is significant for *ACHS*, this means that the surface of the convex hulls increases more for MVM than for GPA, but that the number of loadings that MVM captures more than GPA, does not increase as a result of it. In other words, MVM does not seem to benefit from the faster increase of the surface of the convex hulls when the number of imputations is increased.

To gain more insight in how the counterintuitive main effect of Pooling technique on the percentage of captured loadings could be explained, the values of the average convex hull surface for MVM were plotted against the values the average convex hull surface for GPA. Figure 2 shows the scatter plot for the 100 replicated data sets of the simulation study, when data are imputed 100 times.

Figure 2 shows that for the lower end of the range, the relationship between the results from GPA and from MVM is near one to one. However, as *ACHS* increases for GPA, MVM has some outlying values which clearly do not follow the linear relationship with GPA. The outliers are probably the reason why on average the surfaces of the convex hulls are higher for MVM than for GPA. A larger surface is more likely to capture the loading of the original sample data than a small one.

7. Discussion

In this paper, Generalized Procrustes analysis (GPA) was proposed for combining the results of PCA in multiple imputation as an alternative to the earlier used Mean Varimax method (MVM) and the Mean correlation matrix (MCM). An advantage of GPA compared to MVM is that it automatically corrects for sign reversals and order changes and also rotates solutions optimally towards each other. MCM does not have the problem of sign reversals and order changes either, but it does not preserve information about the amount of uncertainty due to the missing data.

Because GPA provides transformation matrices, scale factors, and centroids within the same framework (i.e., using one loss function), fit measures can be derived that give additional indications of the amount of uncertainty caused by the missing data, as well as the surfaces of the convex hulls indicating the variability in the loadings due to missing data.

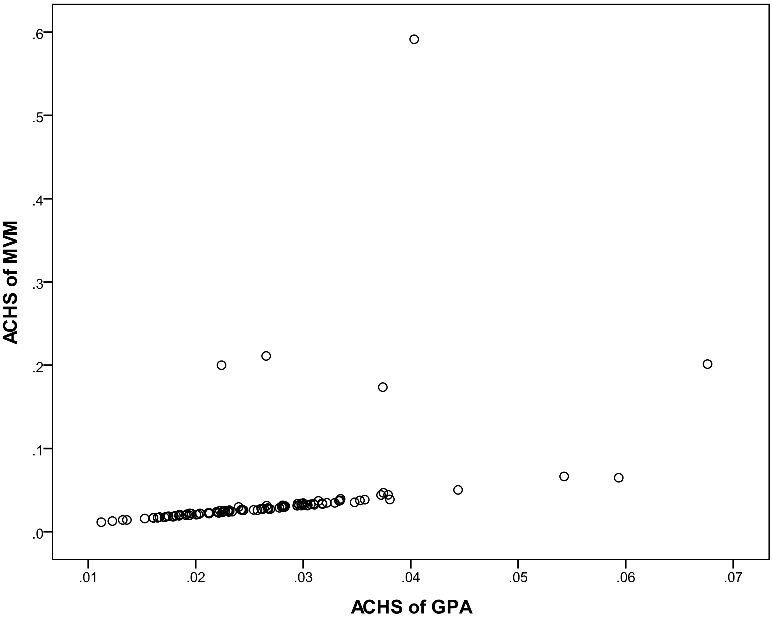

Figure 2. Scatter Plot of ACHS Resulting from GPA (x-axis) Plotted Against ACHS Resulting from MVM (y-axis), Using 100 Imputations.

Such fit measures are discussed by Lingoes and Borg (1978), Gower (1975) and Commandeur (1991). These fit measures can be very useful when an applied researcher wants to draw conclusions about the quality of individual items in a questionnaire. Suppose an item has a small loading in the centroid solution but fit measures and loading plots indicate a high amount of uncertainty of this point estimate. In that case the researcher may conclude that the missing data simply introduces too much uncertainty to conclude that it was unrelated to any of the components, and that more research is needed to draw conclusions about its quality. Without such information the researcher would have concluded that the item has low quality and that it needs to be removed from the questionnaire.

In the first simulation study the performance of GPA was compared with MVM and MCM in three different populations with different factorial structures. Two other factors in this study were sample size and percentage of missingness. This study showed that for MVM sign reversals and order changes of components indeed occurred. For this reason MVM should be avoided.

The only remaining question of the first study is which of the two remaining methods GPA and MCM should be preferred. The root mean squared bias and the mean bias do not provide a clear answer because GPA produced the smallest root mean squared bias, whereas MCM produced the smallest mean bias. However, the authors prefer the GPA approach because this method preserves information about the uncertainty due to missing data, which was the topic of the second simulation study.

In the second study we looked at how GPA represents the uncertainty about the loadings as a result of the missing data. Here, it turned out that 5 imputations are too few to capture a fair amount of the loadings. Simulations were run with 20 and 100 imputations as well, and it turned out that at 100 imputations about 80% of the loadings were captured. To get some impression whether these results generalize to data sets with other properties (different dimensional structures, different number of items), we also looked at the results for five-dimensional data (results not shown). These results indicated that for these data sets about 80% of the loadings were captured as well. Thus, for now we can give a rough guideline that for 15% missing data 100 imputations should be used if the researcher wants to capture about 80% of the loadings using the convex hulls. However, more research on this is still needed.

The second study also showed that when convex hulls are computed using MVM, they capture slightly more loadings than convex hulls resulting from GPA do. However, it turned out that in some individual cases MVM produced substantially larger convex hulls than GPA did in the same data set due to sign reversals that were discussed before. Since larger convex hulls are also more likely to capture loadings of the original sample data, this larger coverage has probably little meaning and could be seen as an artifact. However, to show that this really is an artifact one would have to impute an infinite number of times. If for an infinite number of imputations the average percentage of captured loadings is 100% for GPA and the average convex hull surface is smaller for GPA than for MVM, this means that GPA is superior to MVM. If, on the other hand, the average percentage of captured loadings of GPA is not 100%, then it means that this method produces bias in the loadings.

Unfortunately it is not possible to impute an infinite number of times so that it remains speculative whether the findings of the second study are due to artifacts. However, the superiority of the GPA approach follows directly from its theoretical properties (filtering out variability due to sign reversals, order changes, and rotational freedom) and the theoretical properties of the convex hulls. Also, note that the purpose of the second study was only to illustrate that the surfaces of the convex hulls produced by MVM suffer from sign reversals, order changes, and rotational freedom, and to investigate whether the convex hulls capture a fair amount of the loadings.

Finally we will address a number of points of discussion. Firstly, in this study the data were imputed using the multivariate normal model

(Schafer 1997, Chap. 5). Although it may safely be assumed that PCA loadings obtained from discrete questionnaire data are unaffected by using this model for multiple imputation, the resulting multiply imputed data sets may not be suited for subsequent analyses for categorical data. It should be kept in mind though that the imputation method used is of less relevance for our comparisons since this study was about investigating the quality of combination techniques used after multiple imputation, not about studying the quality of one or more multiple-imputation methods. However, if analyses for categorical data are to be carried out after PCA, it may be better to either round the imputed values or to use a method that can impute categorical data, such as Predictive Mean Matching (Little 1988; Rubin 1986). In this study it was decided not to round the imputed values because for PCA this is not necessary and because we did not want to introduce additional rounding errors.

Secondly, the differences between methods were extremely small. This raises the question whether higher percentages of missing data could have made the differences between methods more clearly visible. Although not realistic, 30% missingness was also studied for some design cells (results not shown). These results revealed that for 30% missingness differences in mean bias between methods were bigger than for lower percentages, but the relations among methods remained the same. For example, MCM had a slightly lower mean bias than GPA at low percentages of missingness. The same was the case at 30% missingness, only the differences between both methods were somewhat larger. For root mean square bias, the interaction between percentage of missingness (including 30%) and pooling technique was not even significant. It should be mentioned that neither of the above-mentioned effects met Cohen's criteria for effect sizes. In short it may be concluded that when using multiple imputation the influence of missing data on PCA results is small. We think that this is because PCA itself is a very robust technique that directs anomalies to later components.

As a third point of discussion Varimax rotated pooled solutions were studied throughout because we needed a common base for comparison of all proposed methods (MVM, MCM, and GPA). For the comparison of the final two selected methods, GPA and MCM, neither Varimax nor Oblique rotations play a role since the pooling is done for the unrotated loadings, which is not the case for MVM. However, for interpretational reasons the resulting pooled unrotated solutions may both be Varimax and Obliquely rotated.

A fourth point of discussion concerns the uncertainty about the loadings, represented by convex hulls. In multiple imputation it is common to study confidence intervals, since multiple imputation was explicitly invented to correct confidence intervals for the extra uncertainty due to missing data (Rubin 1987, p. 2). In our study we only looked at the uncertainty of the component loadings due to the missing data using convex hulls. The variability so depicted differs from confidence intervals in the usual sense in that they provide an indication about the uncertainty of the *sample* loadings. However, in most cases confidence intervals represent uncertainty about a *population parameter*. This difference in studying variability is inherently related to the fact that in most principal component analyses no distributional assumptions are made, and the technique is used in an exploratory manner. If one desires real confidence intervals for population loadings, one should turn to bootstrap confidence intervals in PCA as discussed in Timmerman, Kiers, and Smilde (2007) which then should also be placed in the context of multiple imputation. See the companion paper by Van Ginkel and Kiers (2011)[1], who also studied the GPA and MCM approach for computing component loadings for multiply imputed data sets, but used this in combination with bootstrap confidence intervals of the population loadings.

Besides bootstrap confidence intervals one can also extend the convex hulls presented here to confidence intervals with specific lower- and upperbounds. They may be constructed using convex hull peeling (Green 1981), or confidence ellipses. Also, see Josse, Pagès, and Husson (2011), who used the latter procedure in a similar context. It should be noted that such confidence intervals differ from regular confidence intervals in that they tell us with a certain degree of certainty between which lower- and upperbounds the *sample loadings* fall (not population loadings), had the data been complete. One condition for using this procedure is that more imputations are used than the number of imputations in this study. For example, 1000 imputations are more appropriate, like bootstrap confidence intervals are usually computed using 1000 bootstrap samples. This makes these approaches inappropriate for smaller numbers of imputations. Convex hulls on the other hand, can be computed at all times.

As a fifth point of discussion it may be wondered whether MVM has had a fair chance in this study. When researchers average loadings of multiply imputed data sets, one may wonder whether sign reversals and order changes are never detected. An alert researcher may notice this anomaly and may decide to change the order of components and reverse signs of loadings manually. We have looked into this option as well (MVM with manual correction for sign reversals and order changes of components) and it is our experience that making these corrections manually is a daunting task. In some ambiguous cases it can be extremely

[1] The research in this paper was conducted after this study but the paper was published earlier.

difficult if not impossible to correct for order- and sign changes using the naked eye.

Alternatively one could consider developing an automated procedure for these corrections using some criterion, and implementing this procedure in statistical software packages. An example of such a criterion is maximizing the sum of congruence coefficients (Tucker 1951). However, a disadvantage of such a procedure is that it only solves the problems of sign reversals and order changes and not the problem of rotational freedom among PCA solutions (see also, Timmerman, Kiers, and Smilde 2007). Given that the GPA approach also solves the rotational problem, it automatically follows that there is little merit implementing MVM with such a correction procedure in statistical software packages.

As a final point of discussion, a possible next step in evaluating the performance of GPA for pooling multiple imputation results is comparing its performance with other methods for dealing with missing data (not necessarily based on multiple imputation). This may be done in a wide variety of contexts such as different missingness mechanisms, different numbers of extracted components, and different strengths of the internal structure of the data set. This will be the topic of future research.

8. Appendix

The artificial populations of the simulation study were created from the two NICHD (1996) data sets and the data from Alisic et al. (2008) in the following way:

1. For each item score in both data sets an error was added using a multinomial distribution with the following properties. Suppose ε_{ij} denotes a random error for item score X_{ij} of person i on item j, then ε_{ij} has the following distribution: $P(\varepsilon_{ij} = -2) = 0.01$, $P(\varepsilon_{ij} = -1) = 0.04$, $P(\varepsilon_{ij} = 0) = 0.90$, $P(\varepsilon_{ij} = 1) = 0.04$, and $P(\varepsilon_{ij} = 2) = 0.01$. If the resulting item score was outside the range of 1-5, the item score was replaced with the closest score within this range. In this way a completely new data set was created from the original data set of the same size, and with similar properties.

2. Step 1 was repeated 999 times. Together with the original data sets, these replications formed three artificial populations of sizes N = 1,016,000, N = 1,278,000, and N = 1,515,000, denoted Population 1 for the Child-Parent Relation Scale data, Population 2 for the Self Scale data, and Population 3 for the KIDSCREEN-27 data, respectively.

The idea behind this procedure is that respondents are 90% confident of their given answers and that only in rare occasions they will choose either one of the adjacent answer categories (8% of the times) or the answer categories above or below the adjacent answer categories (2% of the times). By repeating this a large number of times, simulated respondents are created who are similar to the respondents in the data set, with respect to response behavior. Also, see Kroonenberg (1983, Chap. 2) who used a similar procedure for simulating data.

References

ACOCK, A., and MARTIN, J.D. (1974), "The Undermeasurement Controversy: Should Ordinal Data be Treated as Interval?", *Sociology and Social Research, 58,* 427–433.

ALISIC, E., VAN DER SCHOOT, T.A.W, VAN GINKEL, J.R., and KLEBER, R.J. (2008), "Looking Beyond PTSD in Children: Posttraumatic Stress Reactions, Posttraumatic Growth, and Quality of Life", *Journal of Clinical Psychiatry, 69*, 1455–1461.

ANDERSON, T.W. (1963), "Asymptotic Theory for Principal Component Analysis", *Annals of Mathematical Statistics, 34,* 122–148.

ARCHER, C.O., and JENNRICH, R. I. (1973), "Standard Errors for Rotated Factor Loadings", *Psychometrika, 38,* 581–592.

BAKER, B.O., HARDYCK, C.D., and PETRINOVICH, L.F. (1966), "Weak Measurement vs. Strong Statistics: An Empirical Critique of S. S. Stevens's Proscriptions on Statistics", *Educational and Psychological Measurement, 26,* 291–309.

BENZÉCRI, J.P. (1973), *L'Analyse des Données. 1. La Taxinomie, 2. L'Analyse de Correspondances,* Paris: Dunod.

BERNAARDS, C.A., BELIN, T.R., and SCHAFER, J.L. (2007), "Robustness of a Multivariate Normal Approximation for Imputation of Incomplete Binary Data", *Statistics in Medicine, 26,* 1368–1382.

BERNAARDS, C.A., and SIJTSMA, K. (1999), "Factor Analysis of Multidimensional Polytomous Items Response Data Suffering from Ignorable Item Nonresponse", *Multivariate Behavioral Research, 34,* 277–313.

BERNAARDS, C.A., and SIJTSMA, K. (2000), "Influence of Imputation and EM Methods on Factor Analysis when Item Nonresponse in Questionnaire Data is Nonignorable", *Multivariate Behavioral Research, 35,* 321–364.

BOLLEN, K.A., and BARB, K.H. (1981), "Pearson's R and Coarsely Categorized Measures", *American Sociological Review, 46,* 232–239.

CHATTERJEE, S. (1984), "Variance Estimation in Factor Analysis: An Application of the Bootstrap", *British Journal of Mathematical and Statistical Psychology, 37,* 252–262.

COHEN, J. (1988), *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.), Hillsdale, NJ: Lawrence Erlbaum Associates.

COMMANDEUR, J.J.F. (1991), *Matching Configurations*, Leiden, The Netherlands: DSWO Press.

COMREY, A.L., and LEE, H.B. (1992), *A First Course in Factor Analysis* (2nd ed.), Hillsdale, NJ: Lawrence Erlbaum Associates.

COSTA, P.T., and MCCRAE, R.R. (1985), *The NEO Personality Inventory Manual*, Odessa, Florida: Psychological Assessment Resources Inc.

D'AUBIGNY, G. (2004), "Une Méthode d'Imputation Multiple, en ACP", paper presented the XXXVIème Journée de Statistique. Montpellier, France, May 2004.

DOERING, T.R., and RAYMOND, H. (1979), "Measurement and Statistics: The Ordinal-Interval Controversy and Geography", *Area, 11,* 237–243.

GIRSHICK, M.A. (1939), "On the Sampling Theory of Roots of Determinantal Equations", *Annals of Mathematical Statistics, 10,* 203–224.

GOWER, J.C. (1971), “Statistical Methods of Comparing Different Multivariate Analyses of the Same Data”, in *Mathematics in the Archaeological and Historical Sciences*, eds. F.R. Hodson, D.G. Kendall, and P. Tautu, Edinburgh: Edinburgh Univ. Press, pp. 138–149.

GOWER, J.C. (1975), “Generalized Procrustes Analysis”, *Psychometrika, 40,* 33–51.

GRAHAM, J.W., and SCHAFER, J.L. (1999), “On the Performance of Multiple Imputation for Multivariate Data with Small Sample Size”, in *Statistical Strategies for Small Sample Research*, ed. R. Hoyle, Thousand Oaks CA: Sage, pp. 1–29.

GREEN, B.F. (1952), “An Orthogonal Approximation of an Oblique Structure in Factor Analysis”, *Psychometrika, 17,* 429–440.

GREEN, P.J. (1981), “Peeling Bivariate Data”, in *Interpreting Multivariate Data*, ed. V, Barnett, New York: Wiley, pp. 3–19.

GRUNG, B., and MANNE, R. (1998), “Missing Values in Principal Component Analysis”, *Chemometrics and Intelligent Laboratory Systems, 42,* 125–139.

HO, P., SILVA M.C.M., and HOGG T.A. (2001), “Changes in Colour and Phenolic Composition During the Early Stages of Maturation of Port in Wood, Stainless Steel and Glass”, *Journal of the Science of Food and Agriculture, 81,* 1269–1280.

HOCK, E. (1984), “The Transition to Day Care: Effects of Maternal Separation Anxiety on Infant Adjustment”, in *The Child and the Day Care Setting*, ed. R. Ainslie, New York: Praeger.

JOLLIFFE, I.T. (2002), *Principal Component Analysis* (2nd ed.), New York: Springer.

JOSSE, J., PAGÈS, J., and HUSSON, F. (2011), “Multiple Imputation in PCA”, *Advances in Data Analysis and Classification, 5,* 231–246.

JOSSE, J., HUSSON, F., and PAGÈS, J. (2009), “Gestion des Données Manquantes en Analyse en Composantes Principales”, *Journal de la Société Française de Statistique, 150,* 28–51.

KIERS, H.A.L. (1997), “Weighted Least Squares Fitting Using Ordinary Least Squares Algorithms”, *Psychometrika, 62,* 251–266.

KNAPP. T.R. (1990), “Treating Ordinal Scales as Interval Scales: An attempt to Resolve the Controversy”, *Nursing Research, 39,* 121–123.

KROONENBERG, P.M. (1983), *Three-Mode Principal Component Analysis*, Leiden, The Netherlands: DSWO Press, accessed January, 2013, from http://three-mode.leidenuniv.nl/

KROONENBERG, P.M. (2008), *Applied Multiway Data Analysis*, Hoboken, NJ: Wiley.

LABOVITZ, S. (1967), “Some Observations on Measurement and Statistics”, *Social Forces, 46,* 151–160.

LINGOES, J.C., and BORG, I. (1978), “A Direct Approach to Individual Differences Scaling Using Increasingly Complex Transformations”, *Psychometrika, 43,* 491–519.

LINTING, M., MEULMAN, J.J., GROENEN, P.J.F., and VAN DER KOOIJ, A.J. (2007), “Stability of Nonlinear Principal Components Analysis: An Empirical Study Using the Balanced Bootstrap”, *Psychological Methods, 12,* 359–379.

LITTLE, R.J.A. (1988), “Missing-Data Adjustments in Large Surveys”, *Journal of Business and Economic Statistics, 6,* 287–296.

LITTLE, R.J.A., and RUBIN, D.B. (2002), *Statistical Analysis with Missing Data* (2nd ed.), New York: Wiley.

MARKUS, M.T. (1994), *Bootstrap Confidence Regions in Nonlinear Multivariate Analysis*, Leiden: DSWO Press.

MASI, A.T., ALDAG, J.C., and CHATTERTON, R.T. (2006), “Sex Hormones and Risks of Rheumatoid Arthritis and Developmental or Environmental Influences”, *Annals of the New York Academy of Sciences, 1069,* 223–235.

MEULMAN, J. (1982), *Homogeneity Analysis of Incomplete Data*, Leiden: DSWO Press.

MILAN, L., and WHITTAKER, J. (1995), “Application of the Parametric Bootstrap to

Models that Incorporate a Singular Value Decomposition”, *Applied Statistics, 44,* 31–49.
NANDAKUMAR, R., YU, F., LI, H.H., and STOUT, W.F. (1998), “Assessing Unidimensionality of Polytomous Data”, *Applied Psychological Measurement, 22,* 99–115.
NICHD EARLY CHILDCARE RESEARCH NETWORK (1996), “Characteristics of Infant Childcare: Factors Contributing to Positive Caregiving”, *Early Childhood Research Quarterly, 11,* 269–306.
OGASAWARA, H. (2000), “Standard Errors of the Principal Component Loadings for Unstandardized and Standardized Variables”, *British Journal of Mathematical and Statistical Psychology, 53,* 155–174.
OGASAWARA, H. (2002), “Concise Formulas for the Standard Errors of Component Loading Estimates”, *Psychometrika, 67,* 289–297.
PIANTA, R.C. (1992), *Child-Parent Relationship Scale*, Charlotsville: University of Virginia.
RADLOFF, L.S. (1977), “The CES-D Scale: A Self-Report Depression Scale for Research in the General Population”, *Applied Psychological Measurement, 1,* 385–401.
RAVENS-SIEBERER, U., AUQUIER, P., ERHART, M., GOSCH, A., RAJMIL, L., BRUIL, J., POWER, M., DUER, W., CLOETTA, B., CZEMY, L., MAZUR, J., CZIMBALMOS, A., TOUNTAS, Y., HAGQUIST, C., KILROE, J, and the EUROPEAN KIDSCREEN GROUP (2007), “The KIDSCREEN-27 for Children and Adolescents: Psychometric Results from a Cross-Cultural Survey in 13 European Countries”, *Quality of Life Research, 16,* 1347–1356.
ROUSSEEUW, P.J., RUTS, I., and TUKEY, J.W. (1999), “The Bagplot: a Bivariate Boxplot”, *The American Statistician, 53,* 382–387.
RUBIN, D.B. (1976), “Inference and Missing Data”, *Biometrika, 63,* 581-592.
RUBIN, D.B. (1986), “Statistical Matching Using File Concatenation with Adjusted Weights and Multiple Imputations”, *Journal of Business and Economic Statistics 4,* 87–94.
RUBIN, D.B. (1987), *Multiple Imputation for Nonresponse in Surveys*, New York: Wiley.
SCHAFER, J.L. (1997), *Analysis of Incomplete Multivariate Data*, London: Chapman and Hall.
SCHAFER, J.L. (1998), *NORM: Version 2.02 for Windows 95/98/NT*, accessed January, 2013, from http://www.stat.psu.edu/~jls/misoftwa.html
S-PLUS 7 for WINDOWS [Computer software], (2007), Seattle, WA: Insightful Corporation.
SPSS INC. (2011), SPSS 19.0 for Windows [Computer software], Chicago: SPSS.
SU, Y.S., GELMAN, A., HILL, J., and YAJIMA, M. (2011), “Multiple Imputation with Diagnostics (mi) in R: Opening Windows into the Black Box”, *Journal of Statistical Software, 45,* 1–31.
TAKANE, Y., and OSHIMA-TAKANE, Y. (2003), “Relationship Between Two Methods for Dealing with Missing Data in Principal Component Analysis”, *Behaviormetrika, 30,* 145–154.
TEN BERGE, J.M.F. (1977), “Orthogonal Procrustes Rotation for Two or More Matrices”, *Psychometrika, 42,* 267–275.
TIMMERMAN, M.E., KIERS, H.A.L., and SMILDE, A.K. (2007), “Estimating Confidence Intervals for Principal Component Loadings: A Comparison Between the Bootstrap and Asymptotic Results”, *British Journal of Mathematical and Statistical Psychology, 60,* 295–314.
TUCKER, L.R. (1951), “A Method for Synthesis of Factor Analysis Studies”, Personnel Research Section Report No. 984, Washington, DC: Department of the Army.
VAN BUUREN, S. (2010), “Item Imputation Without Specifying Scale Structure”, *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences, 6,* 31–36.

VAN BUUREN, S., BRAND, J.P.L., GROOTHUIS-OUDHOORN, C.G.M., and RUBIN, D.B. (2006), "Fully Conditional Specification in Multivariate Imputation", *Journal of Statistical Computation and Simulation, 76,* 1049–1064.

VAN GINKEL, J.R. (2010), "Investigation of Multiple Imputation in Low-Quality Questionnaire Data", *Multivariate Behavioral Research, 45,* 574–598.

VAN GINKEL, J.R., and KIERS, H.A.L. (2011), "Constructing Bootstrap Confidence Intervals for Principal Component Loadings in the Presence of Missing Data: A Multiple-Imputation Approach", *British Journal of Mathematical and Statistical Psychology, 64,* 498–515.

VAN GINKEL J.R., and KROONENBERG, P.M. (2009), "Using Generalized Procrustes Analysis to Combine the Results from Principal Components Analysis in Multiple Imputation", presentation given at the 16th International Meeting of the Psychometric Society, Cambridge, July 2009.

VAN GINKEL, J.R., VAN DER ARK, L.A., SIJTSMA, K., and VERMUNT, J.K. (2007), "Two-Way Imputation: A Bayesian Method for Estimating Missing Scores in Tests and Questionnaires, and an Accurate Approximation", *Computational Statistics and Data Analysis, 51,* 4013–4027.

WEISSTEIN, E.W., "Heron's Formula", MathWorld–A Wolfram Web Resource, accessed, January, 2013 from http://mathworld.wolfram.com/HeronsFormula.html

WENTZELL, P.D., ANDREWS, D.T., HAMILTON, D.C., FABER, K., and KOWALSKI, B.R. (1997), "Maximum Likelihood Principal Component Analysis", *Journal of Chemometrics, 11,* 339–366.

YUAN, Y.C. (2011), "Multiple Imputation using SAS Software", *Journal of Statistical Software, 45,* 1–25.

ZUCCOLOTTO, P. (2008), "A Symbolic Data Approach for Missing Values Treatment in Principal Components Analysis," *Statistica Applicazioni, 6,* 153–180.