A comparison of confirmatory factor analysis methods
Stuive, Ilse

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2007

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):
8.
The Common Multiple Group Method

8.1. Introduction

In Chapters 5 and 6 the OMG and CCF methods were extensively compared. In Chapter 5 it was shown that in general the OMG method outperformed the CCF method with the detection of incorrect assignments. Furthermore, the OMG method also performed rather well with the detection of correct assignments when the amount of unique variance was equal to or lower than 49%. Unfortunately, the OMG method was less able to detect the correct assignments when the amount of unique variance was fixed at 81%. The CCF method appeared to outperform the OMG method in this condition. An explanation of this finding is that the OMG method does not take into account an estimate of the amount of unique variance whereas the CCF method does. This could especially be a problem when the amount of unique variance is higher than 49%. Because both methods seem to complement each other in case of high amounts of unique variance we decided to study a method that seems to have gone largely unnoticed, but that interestingly shares properties of both the CCF and OMG method: the so-called Common Multiple Group method (Guttman, 1945). Computationally, the Common Multiple Group (CMG) method is highly similar to the OMG method discussed in Chapter 2 with the exception that, with the CMG method, so-called communalities have to be estimated. These communality estimates are placed on the diagonal of the correlation matrix used with the standard OMG analysis. A communality estimate is an estimate of the amount of common variance explained by an item and is equal to one minus its unique variance. This shows its similarity with the CCF method where the amount of unique variance is taken into account by the common factor model. The CMG method and procedures for estimating the communalities are described in Section 8.2.

The main goal of this chapter is to evaluate the performance of the CMG method, and compare it with the results of the OMG and CCF methods presented in Chapters 5 and 6. For an optimal comparison we performed two separate studies where the CMG method will be used on the same data sets as used in Chapters 5 and 6, respectively. In Study 1 we will evaluate how well the CMG method is able to detect correct and incorrect assignments of items to subtests. The obtained results will be compared with those of the OMG method as well as the CCF method. For the CCF
method we decided to choose the RMSEA > .05 & SRMR > .06 criterion as it was shown to be the overall best CCF criterion in Chapter 5. It was expected that the CMG method, combining properties of both the OMG and CCF method, might give high proportions of correct judgments also with high amounts of unique variance, with both correct as well as incorrect assignments of items to subtests.

In Chapter 6 we evaluated several adjustment procedures for both the OMG and CCF method. It was studied how well both methods were able to give correct suggestions to adjust an incorrect assignment of items to subtests in such a way that hopefully the correct assignment was obtained. As it is not only important for the CMG method to detect incorrect assignments, but also to give correct suggestions in these situations, we decided to compare several CMG adjustment procedures with three adjustment procedures presented in Chapter 6: one-step OMG, iterative OMG and the CCF method procedure that performed best (i.e., the one only allowing modifications in $\Lambda$, using no stopping criterion). As the iterative OMG method outperformed one-step OMG, also an iterative procedure of the CMG method will be suggested and evaluated in this chapter.

The main focus of this chapter will be on the question of whether the CMG method performs better than the OMG method especially in those conditions where the amounts of unique variance are rather high. Therefore, we will only focus on those effects discussed in Chapters 5 and 6 that were related to the influence of the amount of unique variance.

Before presenting the results of the two comparative studies the CMG method will be discussed in more detail.

### 8.2. The CMG method

The CMG method (Guttman, 1945) uses the common factor model mentioned in Section 3.1. With this model it is defined that each item score consists of a common and a unique part. Correlations between items are due to the fact that they share an underlying common part at least to some extent. Therefore, the common parts are also known as the correlation producing parts of the items. The unique part of an item consists of those aspects that are unique for a specific item, such as measurement error, and is therefore assumed to be independent of the common and unique parts of the other items.

To perform the CMG method, we again need to compute the correlation matrix of the item scores, summarized in matrix $X$. When we take into account that
the standardized item scores (collected in $Z$) consist of a common part $Z_c$ and a unique part $Z_u$ and that those two parts are independent, we can write the correlation matrix $R$ follows

$$ R = \frac{1}{n} Z'Z = \frac{1}{n} \left( (Z_c + Z_u)'(Z_c + Z_u) \right) $$

$$ = \frac{1}{n} Z'_cZ_c + \frac{1}{n} Z'_uZ_u + \frac{1}{n} Z'_cZ_u + \frac{1}{n} Z'_uZ_c = \frac{1}{n} Z'_cZ_c + \frac{1}{n} Z'_uZ_u $$  \quad (8.1)

From (8.1) we see that the correlation matrix $R$ can be written as a function of a common part $C = \frac{1}{n} Z'_cZ_c$ and a unique part $U = \frac{1}{n} Z'_uZ_u$.

As the unique parts of the items are assumed to be uncorrelated, matrix $U$ is a diagonal matrix with the unique variances of the items on the diagonal. Because the off-diagonal elements in $U$ are zero, the off-diagonal elements in $C$ are identical to those in $R$. This is in accordance with the fact that the common part is the correlation producing part of the items. The diagonal elements in $C$ are now equal to the diagonal elements in $R$, which are equal to one, minus the unique variances of the items on the diagonal of $U$. These diagonal elements in $C$ are known as the communalities of the items. The CMG method is distinguished from the OMG method by using matrix $C$ instead of $R$.

In Section 3.1 we already mentioned that the unique variances, and thus also the communalities, are unknown properties of the items. The communalities therefore have to be estimated. Several procedures have been suggested to estimate these communalities, e.g., see Gorsuch (1983, pp. 103 – 109). In the simulation studies discussed in this chapter we chose to use two simple estimation procedures suggested by Gorsuch and a more complex estimation procedure suggested by Ten Berge and Kiers (1991). The first two estimation procedures suggested by Gorsuch are directly based on the correlation matrix $R$. With the CMG method, Gorsuch (1983, p. 87) suggested only to use a part of this matrix $R$ to estimate the communality for a specific item, namely only the correlations between the items assigned to the same subtest should be taken into account. This implies that possible communalities of an item with items not belonging to the same subtest are ignored in the estimation. We will indicate this by using $R_{\text{subtest}}$ instead of $R$.

First, we chose to use the squared multiple correlation (smc, also see Nunnally, 1978, p.410) as a communality estimate. This estimate is considered to be a lower bound for communality in the population (Guttman, 1956). For each $j$th item, the squared multiple correlation $R_{j}^2$ is computed by
where \( r_{jj} \) is the diagonal element of the inverse of the correlation matrix \( \mathbf{R}_{\text{subtest}} \) corresponding to item \( j, j=1, \ldots, m \). These \( m \) estimates are placed on the diagonal of \( \mathbf{C} \). This method will be abbreviated by CMGsmc.

Secondly, we chose to use the “highest correlation” procedure, also suggested by Gorsuch (1983). With this procedure the absolute highest correlation of an item with another item in \( \mathbf{R}_{\text{subtest}} \) is chosen as the communality estimate of that specific item. This method will be abbreviated by CMGhigh.

The third estimation procedure, known as Minimum Rank Factor Analysis (MRFA; Ten Berge & Kiers, 1991), is not directly based on \( \mathbf{R} \). This method is based on minimizing the nonexplained common variance \( \mathbf{R} - \mathbf{U} \), where \( \mathbf{U} \) is a diagonal matrix containing the unique variances as indicated above, given a fixed number of major common factors. Here this number was chosen equal to the number of specified subtests \( q \). MRFA determines the diagonal elements in \( \mathbf{U} \) in such a way that the nonexplained common variance is minimized using \( q \) common factors. The diagonal elements in \( \mathbf{R} - \mathbf{U} \) then contain the communality estimates. This method will be denoted as the CMGmrfa.

In Study 1 we also decided to use the population communality values as estimates of the communalities in the samples. Of course in practice these values are not known. However, as we constructed the population data, we have the advantage of knowing these communality values. As with increasing sample sizes the communality estimates converge to these population values, these population values are the best possible communality estimates. This allows us also to judge the quality of the three communality estimates. Here, the population communality estimates were defined as the sum of the major and the minor variance, given in Table 5.2. This is equal to one minus the amount of unique variance. So, for example, for all combinations with 25% unique variance the communality estimate was equal to 0.75.

After determining matrix \( \mathbf{C} \), the correlations between items and subtests are computed in a way similar to (2.1), namely

\[
\mathbf{S} = \mathbf{C} \cdot \mathbf{W} \cdot \mathbf{D}
\]

where \( \mathbf{C} \) is the \( m \times m \) correlation matrix with communality estimates of the \( m \) items on the diagonal, \( \mathbf{W} \) is a \( m \times q \) weight matrix containing only zeros and ones defining the assignment of items to subtests and \( \mathbf{D} \) is a diagonal matrix of order \( q \times q \) containing the inverses of the subtests’ standard deviations. We chose to use no corrections for self-correlation and test length with the CMG method, because a preliminary simulation study showed that results were clearly worse when corrections were used.
8.3. Study 1: Comparing the ability to detect correct and incorrect assignments using the CMG, OMG and CCF methods

8.3.1. Method Study 1

In Study 1, the CMG method was applied to the same data sets as described in Section 5.2 using the four communality estimation procedures discussed above. It was investigated how well the CMG method was able to indicate the correct and incorrect assignments of items to subtests and its performance was compared with that of the OMG and the best CCF method (i.e., the one with as criterion RMSEA>.05 & SRMR>.06).

As the CMG method only differs from the OMG method by using $C$ instead of $R$ the same rule as used for the OMG method is used to decide whether or not a correct judgment was made by the method. So, the CMG method simply indicates that the assignment is correct when for each item the correlation in $S$ is highest with the subtest to which it was assigned in the true partition and incorrect otherwise.

8.3.2. Results Study 1

In this study the four CMG procedures were applied to the data sets described in Section 5.2. To make the results in this section comparable to those presented in Section 5.3 we decided to discuss the same effects as presented in that section. However, we will restrict the discussion here to those effects including the amount of unique variance because this is the main focus of this chapter. First, the results will be discussed when the correct assignment was used on the data.

Figures 8.1a, b and c show for 25%, 49% and 81% unique variance respectively, the proportion of correct judgments for the CCF, OMG and four CMG methods as a function of sample size.

From Figures 8.1a and 8.1b we see that with 25% and 49% unique variance the best results were obtained using CMGhigh, CMGmrfa and CMGpop. Only with 49% unique variance and sample sizes lower than or equal to 100 the CMGmrfa method clearly outperformed the CMGhigh and CMGpop methods. OMG and CMGsme performed quite well in 25% unique variance conditions and in 49% unique variance conditions for sample sizes of at least 200. CCF performed well only for sample sizes of at least 200.
Figure 8.1a, b and c. Proportion of correct judgments using the CCF, OMG and four CMG methods on data sets consisting of 25% (a), 49% (b) or 81% (c) unique variance, as a function of sample size when a correct assignment was used on the data.
Figure 8.1c shows that when the amount of unique variance was as high as 81%, the CMGmrfa method performed better than the OMG method but still considerably worse than the CCF method. The CMGhigh and CMGpop also performed better than the OMG method but far worse than the CMGmrfa method. The CMGsmc method again performed equally well or worse than the OMG method.

For the case where incorrect assignments were used, in Chapter 5 we found that the OMG method performed perfectly well in all conditions. This was now also found for all three CMG methods, which is therefore not displayed here again. It should be remembered that here CCF performed considerably worse than OMG, and hence also than all CMG variants.

8.4. Study 2: Comparing the ability to adjust incorrect assignments of items to subtests using the CMG, OMG and CCF method

8.4.1. Method Study 2

In this study we used the CMG methods on the same data sets as described in Sections 6.2 and 6.4. The performance of the CMG methods will be compared to the one-step OMG, iterative OMG and the best CCF procedure (only allowing modifications in A and no stopping criterion). We chose to study only the CMGhigh, CMGsmc and CMGmrfa methods as the CMG with population values as communality estimates is not possible to apply in practice and has shown to perform rather similarly to the CMGhigh method in Section 8.3.2. In Chapter 6 we found that the iterative OMG procedure performed best. For this reason we decided also to construct iterative variants of the CMGhigh, CMGsmc and CMGmrfa methods. The iterative CMG variants were created in a similar way as described in Section 6.1.1 for the iterative OMG method, and will be denoted here as the iterative CMGhigh, iterative CMGsmc and iterative CMGmrfa procedures. To emphasize the difference with the iterative procedures, the CMG methods mentioned in the previous section will here be denoted as the one-step CMGhigh, one-step CMGsmc and one-step CMGmrfa procedure. So, in total six CMG procedures will be compared with the two OMG procedures and the best CCF procedure. Again the proportion of correct judgments will be used as the dependent variable under study. As with the OMG method, an obtained simple structure was considered to be correct with the CMG
method when it equalled the simple structure of the true assignment and incorrect otherwise.

8.4.2. Results Study 2

In Chapter 6 the results were subsequently discussed for data sets consisting of four items per subtest (Section 6.3) and six items per subtest (Section 6.5). This will also be done in this Results section. As in Section 8.3.2 we will only focus on those influences discussed in Chapter 6 that included the influence of the amount of unique variance.

In Section 6.3 we already saw that the two OMG adjustment procedures performed perfectly well with 20% and 50% unique variance. Similar results were observed here with the CMG method. So like OMG, CMG outperformed CCF in these conditions, especially for sample sizes of 100 and 200. Therefore, we will only discuss the results when the amount of unique variance was fixed at 80%. The solid bars in Figure 8.5 indicate the one-step procedures and the CCF method. The striped bars indicate iterative OMG and CMG procedures. For both the OMG and CMG method we see that in general the iterative procedures resulted in higher proportions of correct judgments compared to the one-step procedures. We also see from Figure 8.5 that with sample sizes of 100 and 200 the iterative CMGhigh procedure performed best. The iterative OMG procedure slightly outperformed the iterative CMGhigh procedure with a sample size of 400. When a sample size of 1000 was used almost all methods seemed to perform perfectly well except for the CCF and one-step CMGsme procedure. The lowest proportions of correct judgments were consistently obtained using the one-step CMGsme procedure. Its performance was slightly improved when the iterative variant of the method was used. However, then its proportions of correct judgments were still often worse than those obtained using the iterative OMG procedure and the one-step and iterative CMGhigh and CMGmrfa procedures.

When the number of items per subtest was increased up to six items, again proportions of correct judgments equal to one were obtained with 20% and 50% unique variance using the OMG and CMG procedures. This was also the case for the condition where the third incorrect assignment type was used on data sets with 50% unique variance. So again CMG, like OMG, either worked equally well as or better than CCF in these conditions. For this reason we will here only present in detail the
Figure 8.5. Proportion of correct judgments when the nine adjustment procedures were used to judge data sets consisting of items with 80% unique variance as a function of sample size when the number of items per subtest was four.

Figure 8.6. Proportion of correct judgments when the nine adjustment procedures were used to judge data sets consisting of items with 80% unique variance as a function of sample size when the number of items per subtest was six.

results when items were constructed with 80% unique variance. Figure 8.6 shows that with a sample size of 100 the one-step and iterative CMGmrfa procedure performed best. With a sample size of 200 the highest proportions correct were obtained using the CMGhigh method. The iterative and one-step variants of this method performed
rather similarly. With higher samples sizes all methods seemed to perform rather well.

8.5. Discussion

In the two studies presented in this chapter we compared the CMG method with both the OMG and CCF method to see whether this method was better able to detect the correct and incorrect assignments especially in those conditions where the amount of unique variance was rather high. Furthermore, we wanted to see whether the method performed at least equally well as the OMG and CCF method in the conditions with lower amounts of unique variance such that this method possibly could be used instead of the OMG and CCF method. For the same reason it was also important to assess how well the CMG method was able to adjust incorrect assignments of items to subtests.

From the first study presented in this chapter we saw that when the amount of unique variance was as high as 81%, the CMGmrf method was better able to detect the correct assignments than the OMG method, but was still worse than the CCF method. Only with sample sizes as high as 1000 rather high proportions of correct judgments were obtained using the CMGmrf method to detect correct assignments with 81% unique variance. However, considering the judgment of the correct as well as the incorrect assignments, this method seems to perform best on average with 81% unique variance. Furthermore, the CMGmrf method was even better able to detect correct assignments with 25% and 49% unique variance compared to the OMG and CCF methods, especially on data sets with rather low sample sizes. This method is therefore also the preferred method for these conditions. With incorrect assignments, the CMGmrf method performed equally well as the OMG, CMGhigh and CMGsmc methods. So, evaluating all possible combinations of sample size and unique variance presented in the first study, the CMGmrf method seems to perform best.

The CMGsmc method always performed worse than or equally well as the CMGhigh method, which indicates that the squared multiple correlation is not the best communality estimate to use with the CMG method. The MRFA communality estimate does seem to be a good estimate to use with the CMG method, not only because the CMGmrf method performed best compared to the other methods, but also because it performed equally well as or even better than the best possible estimation of the population communality, namely the population value itself.
In the second study we evaluated how well several adjustment procedures of the CMG method performed in comparison to OMG and CCF adjustment procedures when an incorrect assignment was used on constructed data sets. In this study also iterative procedures were introduced for the three CMG methods. With 20% and 50% unique variance the CMG procedures performed perfectly well and equal to the iterative OMG procedure. In Section 6.3 we already saw that in these conditions the CCF method performed slightly worse compared to the iterative OMG procedure when sample sizes used were lower than 400. So, in these conditions both the iterative OMG and CMG procedures seem to be preferred over the CCF procedure. With 80% unique variance the iterative CMGhigh procedure performed best. The CMGmrfa method often performed slightly worse than the CMGhigh method, but better than the CMGsmc method, and often better than the iterative OMG method.

In general an iterative adjustment procedure seems to perform better than a one-step procedure for both the OMG and CMG methods when incorrect assignments of items to subtests are used. This can also be the case when a correct assignment is used. It is possible that the one-step procedure inadvertently judges a correct assignment as incorrect, whereas an iterative procedure after an initial wrong adjustment converges to a solution that indicates that the assignment used is correct. To see whether an iterative procedure also improves the detection of correct assignments we used the one-step and iterative OMG and CMG procedures on the data sets from Study 2 with the distinction that now the correct assignment was used. From this study it appeared that the detection of correct assignments only improved considerably with the CMGsmc method but not to such an extent that it outperformed the CMGmrfa method.

Overall, the CMG method seems to be a promising method combining aspects of both the OMG and CCF method. Combining the results from Study 1 and Study 2, CMGmrfa seems to be the best CMG method. It performs well in judging correct as well as incorrect assignments and is often able to give correct suggestions to adjust incorrect assignments. The CMGmrfa method is also easy to interpret which is a property that it shares with the OMG method. The only disadvantage is that it is not yet easy to implement this method in often used statistical packages such as SPSS.