A comparison of OMG variants

4.1. Introduction

In Chapter 2 a thorough description of the OMG method was given. We mentioned that the OMG method is based on the computation of correlations between items and subtests, where subtests are defined as simple sums of several items. In Section 2.2 two sources were discussed that can cause spuriously high correlations between items and subtests, namely self-correlation and test length. Self-correlation occurs when an item is correlated with a subtest to which it is assigned by the researcher. Then, the item is partly correlated with itself, which inflates the item-subtest correlation artificially. We showed that we can correct the item-subtest correlation for self-correlation by temporarily removing that specific item from the subtest with which it is correlated. Furthermore, we mentioned that differences in subtest length could cause differences in reliability and therefore differences in the height of correlations involving the subtest. We introduced formula (2.1) which can be used to correct correlations for test length differences.

Both corrections seem necessary on theoretical grounds. However, an important question is whether they actually do so in practice. Do the corrections increase the number of correct decisions made using the OMG method and if so, under which conditions are they most useful? It could, for example, be expected that a correction for self-correlation becomes less influential to the outcome when the number of items per subtest increases. This is because the effect of the perfect correlation on the item-subtest correlation becomes smaller with an increasing number of items per subtest (Nunnally, 1978). Furthermore, it is also plausible that the test length correction is mainly important when the numbers of items per subtest differ a lot. Thus it can be the case that the corrections are necessary on theoretical grounds but that ultimate conclusions do not change or, worse, that more incorrect conclusions are drawn when using corrected correlations. It is, for example, possible that two item-subtest correlations for one item are rather similar before any correction is performed and that after a correction for self-correlation the correct item-subtest correlation is lowered to such an extent that the item is incorrectly assigned to the other subtest.

The first goal of this chapter is to study the influence of these corrections on the number of correct conclusions that are drawn when the OMG method is used on
several simulated data sets. The simulated data sets will be constructed such that several properties of the data sets are varied. This allows us to relate differences in conclusions to differences in characteristics of the data. This study will be discussed in Section 4.2.

The second goal is to study the necessity of standardizing the items before performing an OMG analysis, which will be done in Section 4.3. Holzinger (1944) introduced the OMG method as a method of factoring a correlation matrix, which implies using standardized items. Standardization will especially be influential when item variances differ a lot across the different items used in a test. When item variances are rather similar the influence of standardization on the ultimate conclusions is expected to be minimal. In these conditions, the use of raw item scores is probably also a good approach. Again, a simulation study will be performed to study the influence of standardizing the items. Under several conditions the influence of standardization will be studied to see if, and under which conditions, standardization can be useful.

We decided to discuss both goals separately as the design would become too complex when all different conditions should be taken into account at once. Furthermore, a small simulation study did not reveal an interaction effect between the influence of the different corrections and standardization.

4.2. The influence of correcting for spuriously high correlations

In this section it will be evaluated whether correcting for spuriously high correlations changes the conclusions drawn from an OMG analysis and, if so, under which conditions. Specifically, we would like to know whether correcting for spuriously high correlations increases the number of correct conclusions. To be able to answer these questions we need to construct population data with a known assignment of items to subtests from which we draw samples. So, for example, we construct population data consisting of 12 items and create the data such that three different subtests can be distinguished by three different, non-overlapping sets of four items. From these population data sets we draw samples on which we perform the different OMG analyses. Before the OMG analyses can be performed on the samples, an assignment of items to subtests has to be specified. When this assignment exactly corresponds to the assignment used to construct the population data, the OMG variants should indicate that the specified assignment is correct. However, when this assignment deviates to some extent from the assignment used to construct the data,
the OMG variants should indicate that the assignment used is incorrect and also suggest how to adjust the assignment in such a way that the correct assignment is obtained.

As we already mentioned in Section 2.1, the assignment suggested by the OMG analysis consists of assigning each item to that subtests with which it correlates highest (after possible corrections of these correlations). So, a correct judgment is made when the suggested assignment of items to subtests matches the assignment used to construct the population data. The proportion of correct judgments is used here as the dependent variable under study.

Four OMG variants were compared on their proportion of correct judgments under varying conditions. With the first OMG variant no correction for spuriously high correlations was performed; this was therefore denoted as the “No correction” OMG variant. The second variant is denoted as the “Test length” variant because here only a correction for test length was performed. The third OMG variant only corrects for self-correlation and is therefore denoted as the “Self-correlation” OMG variant. Finally, the “Both corrections” OMG variant denotes the variant correcting for both self-correlation and test length.

4.2.1. Method

To study whether the different OMG variants resulted in different proportions of correct judgments, population data were constructed with a specific assignment of items to subtests and from these populations samples were drawn. The common factor model presented in (3.1) was used to construct the population data, which is for convenience repeated below:

\[ x = \Lambda \xi + \delta \]  

(3.1)

Matrix \( \Lambda \) is a simple structure loading matrix, containing only zeros and ones. The ones in \( \Lambda \) indicate the assignment of items (rows) to subtests (columns) in the population. In the current study, a \( \Lambda \) with three columns was used. The number of items per subtests was varied such that seven different assignments were specified. In the first four assignments each subtest was indicated by 3, 4, 6 or 8 items per subtest, respectively. Characteristic for the other three assignments was that each subtest was indicated by a different number of items. In the fifth assignment, the three subtests were indicated by 3, 4 and 5 items, respectively. With the sixth assignment, 5, 7 and 8 items were used for each subtest and with the seventh assignment the number of items per subtests was 3, 6 and 9, respectively. The factor scores in \( \xi \) were drawn at random.
from a multivariate normal distribution $N(0, \Phi)$, where $\Phi$ is a prespecified correlation matrix of the three subtests. The correlations between the three subtests were taken to be equal, and across conditions varied to be 0, 0.3 or 0.5. The random unique scores were drawn from a multivariate normal distribution $N(0, \Theta)$, where matrix $\Theta$ is a diagonal matrix of order $m \times m$ with on the diagonal the unique variances of the $m$ items.

The unique variances and sample sizes were varied. Namely, the unique variances were varied such that 20%, 50% or 80% of the total amount of variance explained by the items was accounted for by the unique part $\delta$ and the sample sizes were taken equal to 50, 100, 200, 400 or 1000. The design was completely crossed and replicated 100 times so that the total number of data sets analyzed was 100 (replications) $\times$ 7 (number of items per subtests) $\times$ 3 (inter-factor correlation) $\times$ 3 (amount of unique variance) $\times$ 5 (sample size) = 31500 data sets, which were generated using MATLAB7 (MathWorks, 2004). The four OMG variants were applied to the resulting 31500 different data sets, which were standardized before analysis.

To each data set we applied the OMG variants specifying the correct assignment and two types of incorrect assignments. The incorrect assignments were obtained from the population assignments by

1. moving 1 item from the first to the second subtest (Type I)
2. moving 1 item from the first to the second subtest and one from the second to the first (Type II)

This resulted in a total of $31500 \times 3$ (assignment) = 94500 analyses with all four OMG variants. The four OMG variants were programmed in MATLAB7 (2004).

4.2.2. Results

To select the independent variables that were most influential to the association of the choice of method with the proportions of correct judgments, we inspected the values of the partial $\eta^2$, resulting from a repeated measures analysis of variance (RMANOVA), as a measure of effect size (e.g., Bakeman, 2005). Because comparing values of this effect size statistic does not rely on statistical assumptions, its value can be used to compare the relative influence of the different variables even when used with binary data.

As the main focus of this study was on the differences between the four OMG variants we specified a within factor, “Method” distinguishing the four variants. Only those partial $\eta^2$ values resulting from the RMANOVA were inspected that included this “Method” factor. Four between factors were used: the number of items per
subtests, sample size, the amount of unique variance and the inter-factor correlation. A second within factor “Assignment” was used to distinguish the correct and incorrect specified assignments. Table 4.1 presents the effects with partial \( \eta^2 \) values higher than .10.

Table 4.1. Partial \( \eta^2 \) values resulting from a RMANOVA on the proportions of correct judgments.

<table>
<thead>
<tr>
<th>Effect</th>
<th>Partial ( \eta^2 ) value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Method × Sample size</td>
<td>0.274</td>
</tr>
<tr>
<td>Method × Sample size × Unique variance</td>
<td>0.338</td>
</tr>
<tr>
<td>Assignment × Method</td>
<td>0.266</td>
</tr>
<tr>
<td>Assignment × Method × Unique variance</td>
<td>0.277</td>
</tr>
</tbody>
</table>

From Table 4.1, we see that sample size, the amount of unique variance and “Assignment” are influential to differences between the proportions of correct judgments using the four OMG variants. We show the influence of sample size, amount of unique variance and their interaction on the proportion of correct judgments for the four OMG variants separately for the correct and incorrect specified assignments in Figures 4.1 and 4.2, respectively. These two figures display all influential effects given in Table 4.1.

Figure 4.1a, b and c show for 20%, 50% and 80% unique variance, respectively the proportion of correct judgments for all four OMG variants as a function of sample size when correct assignments were used. Around all points in Figure 4.1, and all other figures presented in this chapter, 95% confidence intervals were plotted. From these figures we see that differences between the four OMG variants increase with increasing amounts of unique variance. With 20% and 50% unique variance, similar and very high proportions correct were obtained with the four different OMG variants for all sample sizes. Figure 4.1c shows that with 80% unique variance, the “No correction” variant performed best (with N=50) and equally well as the “Test length” variant with sample sizes larger than 50. With a sample of size 1000 all variants performed very well.

Rather different results were obtained when incorrect assignments were specified and tested on the data with the four OMG variants, see Figures 4.2. From Figure 4.2a we see that again the four OMG variants performed equally well with 20% unique variance independent of sample size. When the amount of unique
Figure 4.1. Proportion of correct judgments using the four OMG variants when a correct assignment was specified for data with 20% (a), 50% (b) or 80% (c) unique variance respectively, as a function of sample size.
Figure 4.2. Proportion of correct judgments using the four OMG variants when an incorrect assignment was specified for data with 20% (a), 50% (b) or 80% (c) unique variance respectively, as a function of sample size.
variance increased to 50% and 80%, as shown in Figures 4.2b and c, the “Both corrections” variant outperformed the other three variants for the medium to high sample sizes. In cases of small sample sizes, the use of the “Test length” variant resulted in the highest proportion of correct judgments.

In contrast to our expectations, the influence of the number of items per subtests on the proportions of correct judgments for the four OMG variants was small: the corresponding partial $\eta^2$ of “Method” × “Number of items per subtest” is equal to 0.027. However, because of our expectations we decided to summarize the proportions of correct judgments for the four OMG methods as a function of the numbers of items per subtest in Figure 4.3. From this figure we see a similar pattern of the proportions of correct judgments for the four OMG methods with several numbers of items per subtest conditions. The lowest proportions of correct judgments were frequently obtained using the “No correction” OMG variant, whereas the highest proportions of correct judgments were obtained when the OMG results were corrected for both sources of spuriously high correlations. These differences seem to disappear with an increasing number of items per subtests. Further inspection of the results revealed that this influence of the number of items per subtests was only observed when using incorrect assignments on the data. With the correct assignments no influence of the number of items per subtests was observed.

![Figure 4.3. Proportion of correct judgments using the four OMG variants as a function of the number of items per subtests.](image-url)
4.2.3. Discussion

In this section we focused on the comparison of four OMG variants to see under which conditions corrections for spuriously high correlations were necessary. From the Results section it can be concluded that especially the amount of unique variance, sample size and assignment type influenced the proportions of correct judgments.

Specifically, all four variants were able to detect correct and incorrect assignments when the amount of unique variance was as low as 20%. With an increase in the amount of unique variance, differences between the four variants increased. Overall, it seems that with 50% unique variance the “Both corrections” variant performed best. With 80% unique variance the “No corrections” variant performed extremely well when correct assignments were used. However, this was accompanied by a very low detection rate of incorrect assignments. When we averaged the average results of the correct assignment conditions with those of the incorrect assignment conditions for the 80% unique variance condition it appeared that with sample sizes lower than 400, none of the four OMG methods was able to judge the assignments correctly more often than in 60% of all data sets. With higher sample sizes the “Both corrections” variant performed best with proportions of correct judgments higher than .9.

In the beginning of this chapter we specified certain expectations about the influence of the number of items per subtest on the proportions of correct judgments for the four variants. This study has shown that its influence seems to be relatively small but clearly present. Furthermore, we expected that a correction for test length would become increasingly important for the proportion of correct judgments when the numbers of items per subtests differed more. This would imply that differences between the “No correction” and “Test length” variant and differences between the “Self-correlation” and “Both corrections” variant would be larger in those conditions where the numbers of items per subtest differed between the three subtests compared to those conditions where the number of items per subtests was fixed across subtests. This expectation was not confirmed in this study. With equal numbers of items per subtest sometimes a slight improvement in the proportion of correct judgments was observed when a correction for test length was performed compared to the “No correction” variant. As in these conditions the test lengths are equal, the correlations are not corrected for their test length but still for their reliability (see Eq. 2.2). So, with equal numbers of items per subtests the “Test length” variant can be considered a correction for reliability.
Overall, we conclude that the “Both corrections” variant resulted in the highest proportion of correct judgments. So, researchers should correct their obtained correlations for both self-correlation and test length to obtain the most reliable conclusions. However, with small samples, e.g., of size 50, it seems preferable to perform no correction at all. In the following sections and chapters we will focus on the “Both corrections” OMG variant. This variant will be called the OMG method in what follows.

4.3. The influence of standardization

In Section 4.2 the four OMG variants were compared using standardized item scores. In (2.1) we showed that the use of standardized item scores corresponds to the use of a correlation matrix. Another possibility is to use the covariance matrix, which means that subtests are created by taking simple sums of the raw item scores. The correlations between items and subtests are then computed using

$$ S = C W D, $$

where $C$ is the $m \times m$ covariance matrix of the $m$ items, $W$ is a $m \times q$ weight matrix containing only zeros and ones defining the assignment of items to subtests and $D$ is a diagonal matrix of order $q \times q$ containing the inverses of the subtests’ standard deviations. $D$ is again used to standardize the subtest scores and thus to obtain correlations in matrix $S$. The only difference with (2.1) is that here the subtest scores are computed using raw item scores instead of standardized item scores.

When the item variances vary, variables with high variances influence the variation of the total score more than variables with a lower variance. When it is considered that variables with high variances are of greater importance than variables with lower variances it is justified to work with raw item scores. However, often all variables are assumed to be of equal importance to the solution. For example, sometimes a questionnaire consists of items where different Likert scales are used. When a questionnaire consists of several items using a five-point Likert scale and some items using a seven-point Likert scale, the items with a seven-point Likert scale are likely to vary more than those with a five-point Likert scale. This is not because of differences in importance of the items or larger variations between individuals, but just because of the fact that a broader scale is used with the seven-point Likert scale. Here, it is very desirable to standardize the items just as in other situations when there is no reason to assume that some items are more important than others.
Sometimes raw item scores are preferable just because of practical reasons and not because of theoretical reasons. Fewer computations are necessary when raw item scores are used and in test construction it is common to compute subtest scores by simply summing the raw scores. Therefore the OMG method using raw scores could sometimes be more practical.

To see whether standardizing the items influences the proportions of correct judgments using the OMG method, again a simulation study was performed. Population data sets were created with varying item variances and OMG on raw item scores was compared to OMG on standardized item scores.

4.3.1. Method

In this section a procedure rather similar to the procedure discussed in 4.2.1 was used. The common factor model was again used to construct population data with a known assignment of items to subtests. In this study, population data for 12 items were constructed where the 12 items were expected to indicate three subtests. Specifically, the first four items were assigned to the first subtest, the second four items to the second subtest and the last four items to the third subtest. From this population data samples were drawn with sample sizes varied to be 50, 100, 200, 400 and 1000. The inter-factor correlations were again varied to be 0, 0.3 and 0.5.

The largest distinction between the present study and the study mentioned in Section 4.2.1 is that here the variances of the items within or between subtests were varied to study the influence of standardization, whereas in the previous study all item variances were created to be equal in the population data sets. So, in the previous study item variances only differed due to sampling variation, whereas here item scores also vary due to differences in population item variances. We decided to create four different variations, where two variations were characterized by item variance differences across and within subtests and two only by item variance differences between subtests. Table 4.2 shows how the item variances were varied in the constructed population data. Between parentheses the standard deviations are also given. Furthermore, the three different subtests are indicated by lines in the Table.

The total item variances given in Table 4.2 were obtained by varying the values in Λ while the amount of unique variance was weighed with one. This resulted in item variances of 1.67, 3.25 and 5.
Table 4.2. Four item variance variations used to manipulate the variances of the items within and between subtests.

<table>
<thead>
<tr>
<th>Item</th>
<th>Variation 1 Variance (Std)</th>
<th>Variation 2 Variance (Std)</th>
<th>Variation 3 Variance (Std)</th>
<th>Variation 4 Variance (Std)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Item 1</td>
<td>1.67 (1.29)</td>
<td>1.67 (1.29)</td>
<td>1.67 (1.29)</td>
<td>1.67 (1.29)</td>
</tr>
<tr>
<td>Item 2</td>
<td>1.67 (1.29)</td>
<td>1.67 (1.29)</td>
<td>1.67 (1.29)</td>
<td>1.67 (1.29)</td>
</tr>
<tr>
<td>Item 3</td>
<td>3.25 (1.80)</td>
<td>3.25 (1.80)</td>
<td>1.67 (1.29)</td>
<td>1.67 (1.29)</td>
</tr>
<tr>
<td>Item 4</td>
<td>5.00 (2.24)</td>
<td>3.25 (1.80)</td>
<td>1.67 (1.29)</td>
<td>1.67 (1.29)</td>
</tr>
<tr>
<td>Item 5</td>
<td>5.00 (2.24)</td>
<td>1.67 (1.29)</td>
<td>3.25 (1.80)</td>
<td>3.25 (1.80)</td>
</tr>
<tr>
<td>Item 6</td>
<td>3.25 (1.80)</td>
<td>1.67 (1.29)</td>
<td>3.25 (1.80)</td>
<td>3.25 (1.80)</td>
</tr>
<tr>
<td>Item 7</td>
<td>3.25 (1.80)</td>
<td>3.25 (1.80)</td>
<td>3.25 (1.80)</td>
<td>3.25 (1.80)</td>
</tr>
<tr>
<td>Item 8</td>
<td>1.67 (1.29)</td>
<td>3.25 (1.80)</td>
<td>3.25 (1.80)</td>
<td>3.25 (1.80)</td>
</tr>
<tr>
<td>Item 9</td>
<td>1.67 (1.29)</td>
<td>1.67 (1.29)</td>
<td>1.67 (1.29)</td>
<td>5.00 (2.24)</td>
</tr>
<tr>
<td>Item 10</td>
<td>3.25 (1.80)</td>
<td>1.67 (1.29)</td>
<td>1.67 (1.29)</td>
<td>5.00 (2.24)</td>
</tr>
<tr>
<td>Item 11</td>
<td>5.00 (2.24)</td>
<td>3.25 (1.80)</td>
<td>1.67 (1.29)</td>
<td>5.00 (2.24)</td>
</tr>
<tr>
<td>Item 12</td>
<td>3.25 (1.80)</td>
<td>3.25 (1.80)</td>
<td>1.67 (1.29)</td>
<td>5.00 (2.24)</td>
</tr>
</tbody>
</table>

In total 5 (sample size) × 4 (variance variation) × 3 (inter-factor correlation) = 60 data sets were created and replicated 1000 times resulting in 60000 data sets. These data sets were analyzed using two OMG variants. With the first OMG variant a correlation matrix was used, whereas with the second variant the covariance matrix was used; both OMG variants used correction for self-correlation and test length. Furthermore, on these sample data sets the same correct and incorrect assignments as in Section 4.2.1 were used to study how well the two OMG variants were able to recognize the assignment used to construct the population data.

4.3.2. Results

The two OMG variants were used to analyze the 60000 different data sets, where we specified one correct and two incorrect assignments. For each analysis it was investigated whether or not the correct assignment was indicated by the OMG method. Again, a correct assignment was said to be indicated when the assignment obtained, by choosing for each item the highest correlation, matched the assignment used to construct the population data.
As in Section 4.2.2 we used RMANOVA to select the independent variables that were most influential to the proportions of correct judgments. Two within variables, “Assignment” and “Method”, were used to cover the three assignment types and the use of a covariance or correlation matrix, respectively. Furthermore, three between variables were used: sample size, variance variation and inter-factor correlation.

Table 4.3. Partial $\eta^2$ values resulting from a RMANOVA on the proportions of correct judgments.

<table>
<thead>
<tr>
<th>Effect</th>
<th>Partial $\eta^2$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Method $\times$ Variance variation</td>
<td>0.016</td>
</tr>
<tr>
<td>Method $\times$ Variance variation $\times$ Sample size</td>
<td>0.013</td>
</tr>
<tr>
<td>Method $\times$ Sample size $\times$ Inter-factor correlation</td>
<td>0.009</td>
</tr>
<tr>
<td>Assignment $\times$ Method $\times$ Variance variation</td>
<td>0.017</td>
</tr>
<tr>
<td>Assignment $\times$ Method $\times$ Variance variation $\times$ Sample size</td>
<td>0.009</td>
</tr>
</tbody>
</table>

Again we decided only to inspect those partial $\eta^2$ values resulting from the RMANOVA containing the “Methods” variable as our main interest is in the difference between using a correlation or covariance matrix. It appeared that overall rather low partial $\eta^2$ values were obtained. We decided only to study those effects with partial $\eta^2$ values near 0.010 or higher. All remaining effects had partial $\eta^2$ values lower than or equal to .006. To cover most influential effects presented in Table 4.3, we decided to present in Figures 4.4a and 4.4b, for the correct and incorrect assignment respectively, the proportions of correct judgments for the two OMG variants as a function of the variance variation and sample size.

Figure 4.4a shows that when a correct assignment was used on the data virtually no differences were observed between the two OMG variants. From Figure 4.4b we see that small differences between the two OMG variants were observed when an incorrect assignment was used on the data. With the first and second variance variation, where the variances differed within subtests, and sample sizes lower than or equal to 100, proportions of correct judgments were slightly higher when the covariance matrix was used. However, opposite results were found with the third and fourth variance variations, where the variances differed across subtests only. Then the analysis of the correlation matrix seemed to be preferable for almost all sample sizes.
Figure 4.4. The proportion of correct judgments when the two OMG variants were used with correct (4.4a) and incorrect assignments (4.4b), as a function of sample size and variance variation.

Because the inter-factor correlation also seemed influential, we studied the results presented in Figure 4.4 for the three different inter-factor correlations separately. We decided not to show these results as for all three inter-factor correlation conditions the differences between the two OMG variants appeared to be rather similar to those presented in Figure 4.4.
4.3.3. Discussion

In this section two OMG variants were compared to see whether standardizing the items influenced the performance of the OMG method when item variances differed considerably across and within subtests. Results have shown that the differences between the two OMG variants are often very small. When a correct assignment was used on the data, virtually no differences between the two OMG variants were found. With a correct assignment, the correlations between items and subtests are probably high enough to compensate for the influence of different item variances.

When an incorrect assignment was used, small differences between the two OMG variants were observed. When item variances varied within and between subtests (indicated by the first two variance variations), the use of the correlation matrix seemed preferable with low sample sizes. With higher sample sizes both OMG variants performed equally well. On the other hand, the use of the covariance matrix resulted in higher proportions of correct judgments when the item variances varied between and not within subtests (indicated by the last two variance variations). This was observed for almost all sample sizes. From these results we may conclude that differences in item variances mainly influence the correlations between items and subtests when an item is incorrectly assigned to a subtest consisting of items all having a different variance than the variance of the incorrectly assigned item itself.

Overall, the OMG results were slightly better when the correlation matrix was used compared to the covariance matrix. This will therefore be the OMG variant used in the sequel of this dissertation.