A comparison of confirmatory factor analysis methods

Stuive, Ilse

Publication date:
2007

1.

Introduction

In the behavioural sciences, questionnaires or tests are frequently used to study psychological constructs, such as personality traits, attitudes, or a person's intelligence. These questionnaires or tests are necessary as these psychological constructs are not directly observable. Therefore, questionnaires have to be constructed in such a way that they consist of several items that are expected to measure such so-called latent variables. The observed items are considered to be at an interval level even though the item scores are usually at the ordinal level (e.g., attitudes can be measured with items using answer categories such as "always", "often", "seldom" or "never"). Considering the items at the interval level has the advantage that powerful statistical procedures can be used to evaluate the questionnaires. In this dissertation I will focus on analysis methods that assume the items to be on an interval scale. When this assumption is violated (i.e., items are on an ordinal scale) analysis results are an approximation.

Frequently, a questionnaire aims to measure several related constructs. These related constructs are then measured by means of subtests that consist of subsets of items supposed to be indicative for the associated latent constructs. For example, an IQ test consists of several items measuring verbal intelligence but it also contains items measuring numerical intelligence. When the items are constructed it is assumed that they measure a certain construct but this has to be evaluated before a questionnaire can be used in practice. For this purpose, the constructed items are usually administered to a large group of subjects. On the basis of the resulting data, test constructers want to verify whether the items can be partitioned into (often prespecified) subtests and if so, whether these subtests are likely to measure the constructs of interest.

One way to obtain insight in possible partitions of the items in subtests is to apply exploratory factor analysis (EFA) to the data. Principal component analysis is the most frequently used EFA method. With PCA, items are summarized into so-called components that consist of linear combinations of the observed items. The first component is obtained by searching for a linear combination of items that explains a maximum amount of variance of the total amount of variance present in the data. In practice, not all of the total variance can be explained by the first component and a second component is obtained that is uncorrelated with the first component and explains as much of the residual variance (total variance minus the variance accounted for by the first component) as possible. This procedure continues until the amount of residual variance explained by a component is small compared to what is explained by previously obtained components. In an EFA, researchers have to specify the number of components. This number is based on issues regarding content and/or data based criteria (like the scree test). Frequently, these obtained components are not easy to interpret as the components are just weighted sums of the items, where weights are chosen in such a way that the maximum amount of variance is explained and not to optimize the interpretability. A solution is to rotate the PCA solution in such a way that better interpretable results are obtained. This rotated solution can consist of orthogonal as well as oblique (correlated) components. Characteristic of these components is that they explain the same amount of variance as the components obtained with PCA but are usually much easier to interpret. In general, the goal is to rotate the solution optimally following a specific criterion. For example, with a Varimax rotation the solution is rotated in such a way that hopefully a simple structure is obtained. Such a simple structure is obtained when each item tends to load high on one component and low on the other components. Next, these obtained components can be interpreted, using the contents of the items that load high on a component, to see whether they are likely to measure a specific underlying construct.

The use of EFA makes sense when researchers only have little theoretical or empirical knowledge about their constructed test, or when researchers are interested in the optimal assignment of items to subtests as far as indicated by the data only. However, researchers often will have certain specific ideas about the assignment of items to subtests and want to verify whether these ideas are supported by the data. It is unlikely that a test is constructed without some a priori idea about the partitioning of items in subtests. Researchers, for example, want to use current theory in their field to make predictions about the assignment of items to subtests. It could also be the case that related empirical studies repeatedly indicated the same assignment of items to subtests. A way to verify whether such an a priori assignment of items to subtests is supported by the data is by using a confirmatory factor analysis (CFA). With CFA it has to be specified in advance how many subtests are expected to be present and which items are to be assigned to which subtest.

Two CFA methods exist to verify whether an a priori assignment of items to subtests is supported by the data, the Oblique Multiple Group (OMG; Holzinger, 1944) method and the Confirmatory Common Factor (CCF; Jöreskog, 1969) method. The CCF method is the most often used method in practice. With the CCF method a model of the data is to be specified using, among other things, the knowledge and ideas about the assignment of items to subtests. Next, the specified model is fitted to the data to see if the a priori assignment of items to subtests is supported by the data.

Several fit indices are provided by popular statistical software programs, such as LISREL (Jöreskog & Sörbom, 2001) to evaluate how well the model is supported by the data.

The CCF method uses the common factor model. Characteristic of this common factor model is that it takes into account the presence of so-called unique factors that are present in the observed scores. These unique factors represent, among other things, the amount of measurement error present in the item scores. This method will be discussed in more detail in Chapter 3.

A second method that can be used to verify an a priori assignment of items to subtests is the OMG method. With this method, to be discussed in Chapter 2, subtest scores are created by taking simple sums of the items that are a priori assigned to the subtest at hand. Next, correlations between items and subtests are computed. Finally, each item is assigned to that subtests with which it correlates highest. The OMG method is a component approach as it uses linear combinations of the items as subtests and does not take into account unique factors present in the data (as the CCF method does).

Surprisingly, the rather simple OMG method is largely ignored in practice while it has much to offer compared to the CCF method. The OMG method is conceptually much simpler than the CCF method and it never fails to find a solution, whereas with the CCF method sometimes no solution is obtained. Furthermore, Nunnally (1978, p.403) noted that “it would be foolish to employ such a complex approach if hypotheses are sufficiently clear that they can be tested more simply and directly by the (oblique) multiple group method”.

The fact that the OMG method is largely ignored in practice could be quite understandable if it would have been shown to perform worse than the CCF method. Researchers probably also tend to think that the CCF method is best as it is performed in the much respected framework of covariance structure analysis (CSA). CSA has become very popular over the years as it can be used to answer a broad range of questions and it allows researchers to obtain estimates of the amount of measurement error present in the observed scores. Furthermore, much research has been performed to optimize the methods within this framework whereas the OMG method has remained untouched after its introduction in 1944. However, surprisingly, two small simulation studies indicated that the OMG method actually performed equally well or even better than the CCF method (Tuerlinckx, Ten Berge & Kiers, 1996; Hendriks & Kiers, 1999). Unfortunately, these two simulation studies were too small to provide a solid basis for practical recommendations.

The goal of the studies described in this dissertation is to provide a thorough comparison of the OMG and CCF method which allows us to make recommendations for use in practice. Using simulated and empirical data, it will be evaluated if and under which conditions the methods differ and which method should be preferred under specific circumstances.

The outline of this dissertation is as follows. First, the OMG and CCF methods are discussed more thoroughly in Chapters 2 and 3, respectively. Both methods are illustrated using an empirical example. In Chapter 4, several choices to be made when using the OMG method are discussed and evaluated. This evaluation results in a best performing OMG variant that will be used as the OMG method in this dissertation. In Chapters 5 and 6, the CCF and OMG methods are compared on the ability to detect correct and incorrect assignments of items to subtests and their ability to adjust incorrect assignments, respectively. In these two chapters the comparisons are mainly based on simulated data sets but also on the empirical data set used to illustrate the methods in Chapters 2 and 3. In Chapter 7 both methods are used to analyse several data sets that are obtained using one specific questionnaire to illustrate their performance more thoroughly on empirical data sets. The results of both methods are compared to see whether similar conclusions are drawn when the two different methods are used on the same empirical data sets. In Chapter 8 an alternative multiple group method is introduced; the so-called Common Multiple Group Method (Guttman, 1945). As its name already suggests, it is a common factor approach that includes an estimate of the unique factor present in the item scores. Therefore, it has some similarities to both the OMG and CCF method. This method is compared with the OMG and CCF method under similar conditions as used in Chapters 5 and 6 to make these results optimally comparable to those presented in these chapters. Finally, in Chapter 9, several OMG procedures will be discussed and evaluated that can be used to handle multiple data sets that are obtained from a single questionnaire. An overall discussion and conclusion is provided in Chapter 10.