Psychological Test and Assessment Modeling, Volume 52, 2010 (3), 298-312

# The c-bifactor model as a tool for the construction of semi-homogeneous upper-level measures

*Karl Schweizer[1], Michael Altmeyer[2], Siegbert Reiß[2] & Michael Schreiner[2]*

**Abstract**

The paper addresses problems resulting from the application of methods, which emphasize homogeneity, in the construction of measures that are expected to represent general upper-level constructs. It distinguishes between homogeneous and semi-homogeneous measures. Whereas homogenous measures allow for one underlying dimension only, semi-homogeneous measures are characterized by the presence of one general and dominating dimension in combination with restricted subordinate dimensions. It is the congeneric model of measurement that tends to create homogeneous measures whereas the c-bifactor model enables the construction of semi-homogeneous measures where the prefix "c" indicates that it is a confirmatory bifactor model. It is made obvious that the successful construction of measures representing upper-level constructs requires the c-bifactor model. The congeneric and c-bifactor models were applied to the social optimism scale since social optimism was known to show a hierarchical structure. As expected, only the c-bifactor model indicated a good model fit.

*Key words*: congeneric model, bifactor model, test construction, homogeneity, social optimism

---

[1] *Correspondence concerning this article should be addressed to:* Karl Schweizer, PhD, Department of Psychology, Goethe University Frankfurt, Mertonstraße 16, 60325 Frankfurt, Germany; email: k.schweizer@psych.uni-frankfurt.de

[2] Goethe University Frankfurt, Frankfurt a. M., Germany

Confirmatory factor analysis (CFA) in the framework of test construction usually aims at the investigation of the homogeneity of the items of a measure. Such an investigation is usually performed on the basis of the congeneric model of measurement (Jöreskog, 1971) that means the consideration of one latent variable in combination with several items serving as manifest variables. The latent variable is assumed to represent the true source of responding to the items. The congeneric model even gave rise to an advanced test theory called congeneric test theory (Lucke, 2005; McDonald, 1999; Raykov, 1997, 2001). However, despite the favorable characteristics and advantages of the congeneric model, an increasing number of deviations from this model can be observed. These deviations can be characterized as bifactor models, a denotation that was introduced quite a time ago (Holzinger & Swineford, 1937) for models that allow items to load on two latent variables. Since the original bifactor model was developed in the context of exploratory factor analysis, it is addressed as *c-bifactor model* in the context of confirmatory factor analysis. There is an accumulation of cases where it is reasonable to assume that more than one latent source influence responding to an item. First, there is the case of a measure imbedded into a multitrait-multimethod design (Campbell and Fiske, 1959). This design gives reason for assuming that traits or abilities in combination with observational methods determine the responses to items. As a consequence, various CFA models allowing each item to load on two latent variables were proposed (e.g., Kenney & Kashy, 1992; Marsh & Bailey, 1991). Second, the effect of different item wordings was repeatedly shown to cause model misfit. The integration of an additional latent variables into the basic scheme provided by the congeneric model of measurement for representing one observational method proved instrumental for achieving good model fit (DiStefano, & Motl, 2006; Rauch, Schweizer, & Moosbrugger, 2007; Vautier, Raufaste, & Cariou, 2003). This provision means the transformation of the standard model of confirmatory factor analysis into a bifactor model. The additional dimension is not apparent in an IRT investigation (Rauch, Schweizer, & Moosbrugger, 2008). Third, the items may show a position effect that is usually the cause of model misfit (Hartig, Hölzel, & Moosbrugger, 2007). In this case there is also the possibility to eliminate model misfit by considering a model of measurement with a second latent variable that accounts for variance due to the position effect (Schweizer, Schreiner, & Gold, 2009). What results is a c-bifactor model.

Finally, it needs to be indicated that the c-bifactor model is not only important for confirmatory factor analysis. It can also replace the congeneric model of measurement as part of a full structural equation model. So the assumption that the responses to items are due to more than one latent source may also be transferred from confirmatory factor analysis to structural equation modeling by integrating the c-bifactor model of measurement.

## The c-bifactor model for representing a hierarchical structure

In this paper another case of deviation from the congeneric model of measurement as the standard model of confirmatory factor analysis is presented and investigated. It is the construction of a measure for an upper-level construct that is part of a hierarchical structure. Various hierarchical structures of personality and ability can be found in psychology literature. Well-known examples are the structures of personality with neuroticism and extraversion-introversion as top-level constructs proposed by Eysenck (1967) and the three-stratum model of cognitive abilities by Carroll (1993). It is a characteristic of such hierarchies that constructs can be found at various levels. In Carroll's model general ability respectively general intelligence is a third-level construct, fluid intelligence a second-level construct and figural reasoning a first-level construct. For an illustration Figure 1 gives a simple two-level hierarchy as CFA model with latent variables for representing constructs and items as manifest variables.

There is the possibility to represent such a hierarchy by means of a corresponding confirmatory factor model (Schweizer, Moosbrugger, & Schermelleh-Engel, 2003). However, in test construction the focus is usually not on the whole but on a specific part of the hierarchy. The constructs assigned to the various levels show different degrees of complexity and, as a consequence, differ according to homogeneity. Lower-level constructs can be expected to be very homogeneous whereas top-level constructs may show an impaired degree of homogeneity. Therefore, confirmatory factor analysis of measure representing one of the low-level constructs can be expected to lead to very favorable results whereas there is no guarantee for success with respect to measures designed for upper-level constructs.

The problem with the homogeneity of measures representing upper-level constructs is that these measures may include items that originate from the same lower-level construct. As a consequence, some items may show considerably higher correlations among each other than with items originating from other lower-level constructs. Figure 2 presents an attempt to make the problem obvious, which results from the latent presence of a lower-level structure.

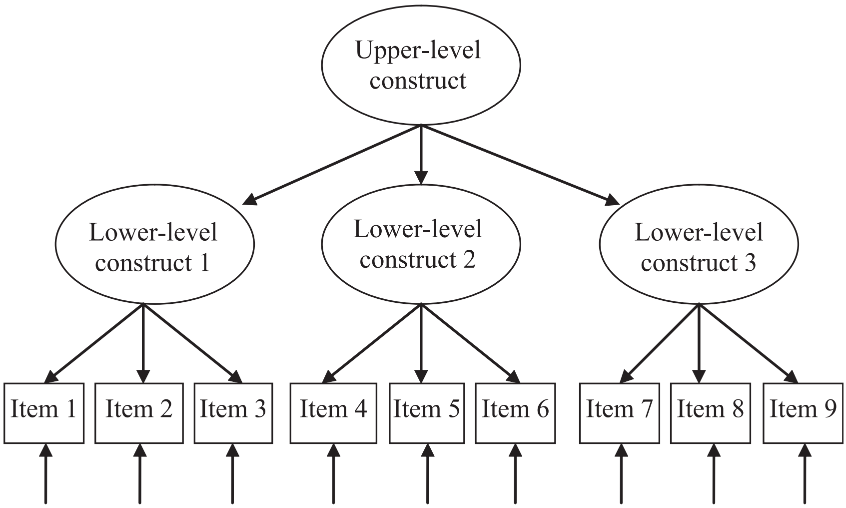


**Figure 1:**
Hierarchical structure with constructs assigned to two latent levels

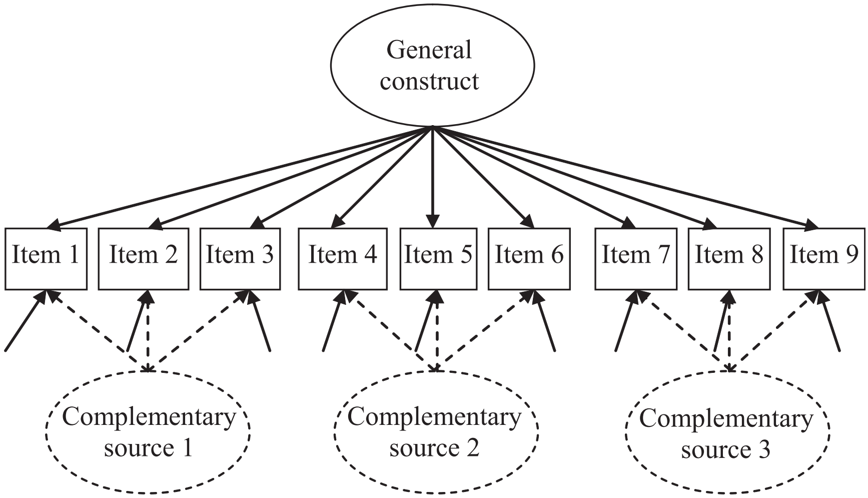


**Figure 2:**
Hierarchical structure presented as general construct with an indication of underlying complementary sources (dashed lines)

In this illustration the presence of lower-level constructs is ignored in that the congeneric model of measurement is assumed. Implicitly the upper-level construct is transformed into a first-order construct denoted general construct, which is to be represented by the measure. The original lower-level constructs are neglected by the model although they are only partly represented by the new first-order construct. However, there are additional components of the original first-order constructs as remainders. These remainders are not automatically integrated into the new first-order construct and can be suspected to be still effective as situational, event-specific or method sources of responding. In order to make the special nature of these additional systematic sources of responding obvious, they are addressed as *complementary constructs*. These additional sources that are neglected are identified by dashed lines. These complementary constructs take the role of disturbances that prevent the achievement of a good model fit in confirmatory factor analysis. In order to avoid confusion at this point, it must be emphasized that there is no exact correspondence between the first-order constructs of Figure 1 and the complementary constructs of Figure 2. There may be some overlap but no correspondence according to the psychological meaning. A similar splitting up into different sources can be found in latent state-trait models (Steyer, Schmitt, & Eid, 1999). Such models require the subdivision of first-order latent variables as seemingly unique sources into state and trait latent variables that are associated with different sources of responding.

The problem resulting from the underlying structure can be solved by selecting an approach similar to the multitrait-multimethod approach (Campbell and Fiske, 1959). This approach demands that systematic sources of responding that can be assumed to be effective as disturbances should be integrated into the research design and the corresponding model of measurement. In the multitrait-multimethod approach the disturbances are mainly observational methods. In the investigation of hierarchical structures the additional constructs play a role that is similar to the role played by observational methods since the additional constructs are sources of a special degree of homogeneity among items. Therefore, it is very reasonable to get the problem into grip by representing the additional constructs in a similar way by corresponding latent variables. Additional constructs as latent variables take the role of method latent variables. As a result of these considerations, a c-bifactor model (cf. Holzinger & Swineford, 1937) is achieved. The illustration of this model is included in Figure 3.

This Figure differs from Figure 2 by the transformation of the dashed lines into solid lines and the change of labels. Since the additional constructs do not fully correspond to the first-order constructs of Figure 1, they are denoted *semi-constructs*. Still it can be assumed that the measure including the first to ninth items represents the general construct if an appropriate degree of model fit is achieved and the sizes of the loadings are acceptable. It does not represent the semi-constructs since these semi-constructs are perceived as sources, which are not common to all of them. No one of them can dominate the responses to all the items. The only way of eliminating the semi-constructs completely would be the elimination of items. If only one item per semi-construct remains, the semi-constructs as underlying constructs are no longer present as a disturbance. However, as a consequence of this provision, the measure may no longer be appropriate because of the low number of items.

The considerations of this section should give the reader an idea of the research problems that may demand the application of the c-bifactor model in order to achieve a reasonable

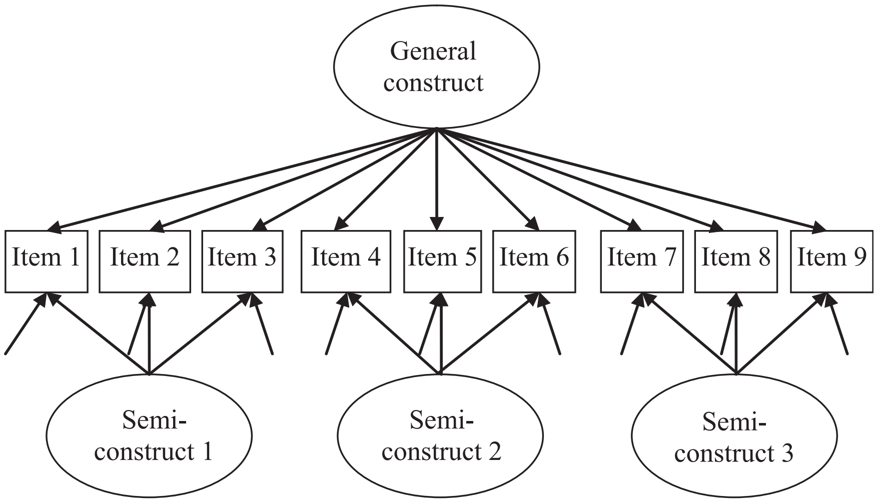


**Figure 3:**
c-bifactor model for a measure representing a general (upper-level) construct in combination with additional semi-constructs

result. But there are also many other research problems that demand the selection of the congeneric model of measurement. So the c-bifactor model simply enlarges the assortment of models, which can be applied for test construction but there is no replacement.

## Formal representation of an upper-level measure

Confirmatory factor analysis usually means selecting the congeneric model (Jöreskog, 1971) for investigating the items of a scale, and the response to an item is considered as the result of contributions of an ability or trait on one hand and error on the other hand. Accordingly, equation (1) defines the response $y_{ij}$ of the $i$th participant to the $j$th item ($j$=1,…,$J$) as

$$y_{ij} = v_j + \lambda_j \eta + \varepsilon_{ij} \ , \qquad (1)$$

where $\nu_j$ characterizes the $j$th item, $\lambda_j$ is the loading of the $j$th item on the latent random variable, $\eta$ the person-specific true score latent random variable and $\varepsilon_{ij}$ the error score random variable associated with the $i$th participant and the $j$th item. The constant $\nu_j$ of the regression model was considered as item easiness (Carmines & McIver, 1981; Ferrando & Lorenzo-Seva, 2005; Jöreskog, 1971; McDonald, 1999) and is usually omitted in the CFA context.

Equation (1) serves well if there is only one source giving rise to the true score. Otherwise it may be necessary to distinguish between parts of the true score due to two or even more sources. However, such a distinction is only reasonable if the sources apply to different sets of items. It is reasonable to assume that one source applies to all the items and, therefore, may be considered as source of a general true score random variable $\eta_{(\text{general})}$ whereas the other ones give rise to a specific true score random variable $\eta_{(\text{specific})k}$ ($k$=1,…,$K$). These latent variables must accompany non-overlapping subsets of items. Their integration into equation (1) gives

$$y_{ij} = v_j + \lambda_{(\text{general})j}\eta_{(\text{general})} + \lambda_{(\text{specific})j}\eta_{(\text{specific})k} + \varepsilon_{ij} \ . \qquad (2)$$

A rather similar extension of the congeneric model was developed for the item factor analysis model (e.g., Bock & Aitkin, 1981; Ferrando, 2005; Gibbons, Bock, Hedeker, Weiss, Segawa, Bhaumik, Kupfer, Frank, Grochocinski, & Stover, 2007). Of course, equation (2) only makes sense if the manifest variables associated with the $j$th specific latent variables are a subset of the manifest variables associated with the general latent variable.

The construction of a measure for representing the upper-level construct of a hierarchy requires the concentration on $\eta_{(\text{general})}$ whereas $\eta_{(\text{specific})}$ is to be considered as a nuisance that eventually must be controlled. Since the effect of $\eta_{(\text{specific})}$ on model fit depends on the types of available items, it is useful to discuss alternative cases. The first case is characterized by the following feature: all the items refer to different subsets. In this case a variant of equation (1) is very suitable for investigating the items:

$$y_{ij} = v_j + \lambda_{(general)j}\eta_{(general)} + \varepsilon_{ij} \tag{3}$$

since $\eta_{(\text{specific})}$ is only of importance if there are at least two items, which are influenced by the same source represented by $\eta_{(\text{specific})}$. The second case is characterized by a mixture of underlying constructs. Some of them are represented by one item whereas others by two or more items. Therefore, the applicability of equation (3) is restricted. Instead a mixture of models of measurement needs to be considered. In some items a model according to equation (2) should serve well whereas in the other items a model according to equation (3) must be preferred. Finally, there is the case that each sub-concept is represented by at least two items. In this alternative the model of measurement according to equation (3) is necessary for all the items. In order to have a general case, it is reasonable to integrate the Kronecker's delta into equation (2) such that

$$y_{ij} = v_j + \lambda_{(\text{general})j}\eta_{(\text{general})} + \delta_k\lambda_{(\text{specific})j}\eta_{(\text{specific})k} + \varepsilon_{ij} \tag{4}$$

and to define $\delta_k$ in such a way that

$$\delta_k = \begin{cases} 1 & if \ g(C_k) > 1 \\ 0 & \text{else} \end{cases} \tag{5}$$

where $C_k$ is the $k$th subset of items referring to $k$th underlying specific construct and $g$ the function that returns the number of items included in $C_k$. Two items as minimum for a homogeneous subset of items may be a good lower limit. However, there may also be reasons for selecting three items as minimum.

The model of measurement characterized by the equations (4) and (5) suggest that the model of measurement may change from one item to the next item. Such a model of measurement considerably increases the flexibility in investigating the items of measures. The c-bifactor model illustrated by Figure 3 can be represented by this model quite well.

## A Comparison of c-bifactor model and hierarchical model concerning test construction

In test construction it is important to have information concerning the appropriateness of items with respect to the construct of interest. Such information is provided by the c-bifactor model, as it is obvious from equation (4). The loadings on the general latent random variable $\lambda_{(\text{general})j}$ ($j$=1,…,$J$) provide information concerning the usefulness of the items for representing the general construct.

In contrast, in the hierarchical model only the loadings on the first-order latent variables, which are specific true score random variables, are available since the second-order latent variable is not directly related to the items. This means that equation (1) must be rewritten in the following way:

$$y_{ij} = v_j + \lambda_{(\text{first\_order})j} \eta_{(\text{first\_order})k} + \varepsilon_{ij} \quad (6)$$

where $y_{ij}$ denotes the response of the $i$th participant to the $j$th item, $\nu_j$ characterizes the $j$th item, $\lambda_j$ is the loading of the $j$th item on the corresponding first-order latent random variable, $\eta_{(\text{first-order})k}$ the $k$th true score first-order latent random variable and $\varepsilon_{ij}$ the error score random variable associated with the $i$th participant and the $j$th item. Since the first-order latent random variable is actually a composite of the second-order latent random variable $\xi$, which is associated with the construct of interest and is accompanied by $\gamma_j$ as weight, and an additional specific latent random variable $\zeta_j$, equation (6) needs to be rewritten as

$$y_{ij} = v_j + \lambda_{(\text{first\_order})j} \left( \gamma_j \xi + \zeta_k \right) + \varepsilon_{ij} \quad . \quad (7)$$

The revised equation also illustrates that the loadings on the first-order latent variable cannot be used for the evaluation of the items.

Furthermore, it makes apparent that the subdivision is really necessary in order to achieve information concerning the appropriateness of an item with respect to the construct of interest. In sum, the c-bifactor model enables the direct estimation of the loadings on the general latent random variable associated with the upper-level construct of interest whereas the hierarchical model makes an indirect estimation necessary.

## Homogeneity and semi-homogeneity and how to keep them separated

*Homogeneity* is a concept that is closely associated with factor analysis. In the spatial notation that presents factors as axes of a high-dimensional space homogeneity means that data can basically be represented by one dimension. So the presence of only one underlying dimension for a set of items characterizes homogeneity. Consequently, it characterizes measures that originate from a factor-analytic investigation in following the basic ideas of factor analysis (Spearman, 1904). In a homogeneous measure all the observations can be best described by this one dimension, and the corresponding responses can be assumed to be due to the same source. Obviously, homogeneity excludes the presence of other dimensions in a set of data obtained by the items of a homogeneous measure. Furthermore, it is quite obvious that homogeneity can easily be achieved in measures representing lower-level constructs whereas it may be difficult to be achieved in upper-level constructs.

The obvious difficulty in achieving homogeneity for upper-level constructs gives reason for proposing another concept, *semi-homogeneity*, since it is not really reasonable to demand homogeneity for corresponding measures. Semi-homogeneity suggests that there is one dominating dimension and does not exclude the restricted presence of other dimensions. There must be one construct giving rise to one dimension that characterizes all the items of a measure and is underlying the responses to all these items. This does not preclude the existence of other constructs that may apply to subsets of items of the measure. So the influence of the other constructs must be restricted to subsets of items so that additional non-dominating dimensions show a clear restriction. A very good example is provided by the specific dimensions resulting from positively and negatively worded items (DiStefano, & Motl, 2006; Rauch, Schweizer, & Moosbrugger, 2007; Vautier, Raufaste, & Cariou, 2003). In this example additional non-dominating dimensions are restricted to the items with positive respectively negative wording.

Measures constructed according to the original version of the congeneric model of measurement can be expected to show a high degree of homogeneity. The emphasis given to one-dimensionality in confirmatory factor analysis becomes especially obvious in considering the model of the covariance matrix and the fit statistics. According to Bollen (1989, p. 35) the model of the covariance matrix is given by

$$\mathbf{\Sigma} = \mathbf{\Lambda}_y \mathbf{\Psi} \mathbf{\Lambda}_y' + \mathbf{\Theta}_\varepsilon \tag{8}$$

where $\mathbf{\Sigma}$ is the $p \times p$ covariance matrix of manifest item random variables, $\Lambda_y$ the $p \times q$ matrix of loadings, $\mathbf{\Psi}$ the $q \times q$ covariance matrix of the latent random variables, and $\mathbf{\Theta}_\varepsilon$ the $p \times p$ diagonal matrix of error variances of latent error variables $\boldsymbol{\varepsilon}$. This model of the covariance matrix is closely linked to the congeneric model so that every change of the congeneric model leads to a corresponding change of the model of the covariance matrix. Actually, the congeneric model is transformed into the other model; it is implicit in the model of the covariance matrix. For example, the inclusion of a specific true-score component into a model of measurement changes the model of the covariance matrix accordingly.

The model of the covariance matrix plays an important role in investigating model fit in confirmatory factor analysis. The majority of major statistics of model fit (SMF) are defined as a function $f$ of the difference between the model of the $p \times p$ covariance matrix $\mathbf{\Sigma}$ and the $p \times p$ matrix of observed variances and covariances $\mathbf{S}$:

$$\mathrm{SMF} = f_{\mathrm{SMF}}\left(\mathbf{S}, \mathbf{\Sigma}\right) \tag{9}$$

(for an overview see Kline, 2005) It is obvious that a bad fit is indicated whenever $\mathbf{\Sigma}$ is constructed for representing one dimension whereas the variances and covariances of $\mathbf{S}$ are due to several dimensions. Therefore, the application of confirmatory factor analysis on the basis of the congeneric model can be expected to produce homogeneity.

Next, the conventional methods of test construction are considered. There is the concept of internal consistency that comes closest to homogeneity. Statistics of internal consistency are applied for evaluating the quality of a measure as a whole. The most well-known and most often applied statistic of internal consistency is Cronbach's (1951) Alpha coefficient. This statistic is usually reported when a measure is constructed according to classical test theory (Gulliksen, 1959; Lord & Novick, 1968). This statistic $\alpha$ reflects all types of covariation among the items:

$$\alpha = \frac{P}{P-1}\left(1 - \frac{\sum_{i=1}^{P} \sigma_i^2}{\sigma_{\mathrm{total}}^2}\right) \tag{10}$$

where $\sigma_i^2$ is the variance of the $i$th item ($i$=1,…,$P$), $\sigma_{\text{total}}^2$ the variance of the item sum and $P$ the number of items. In this formula $\sigma_{\text{total}}^2$ takes a crucial role since its size to a considerable degree reflects the correlations among the items. Furthermore, the size of total variance can be large or low independently of whether the correlations are due to one source or several different sources. Consequently, $\sigma_{\text{total}}^2$ and also $\alpha$ can be large because of the contribution of two or even more underlying dimensions.

Because of the dependence on the correlations it can indicate a high degree of consistency although homogeneity is not given. Therefore, an impairment of homogeneity may not be obvious in Cronbach's Alpha coefficient. Such a lack of sensitivity can be found in methods that concentrate on the investigation of consistency in responding.

## An example

This section reports the investigation of the items of a measure, which is expected to represent an upper-level construct. The measure that serves as an example is taken from optimism research. In 1997 a scale for measuring social optimism was published (Schweizer & Schneider, 1997). The exploratory investigation of the items of this scale revealed a hierarchical structure. Besides the general factor there were lower-level factors that were denoted optimism concerning economic affairs, optimism concerning predicted disasters and optimism concerning the future controllability of violence. Structural investigations revealed a rather complex structure because there were also effects of different item wordings (Schweizer & Rauch, 2008). However, Alpha consistencies and part-whole correlations of an acceptable size were observed for the general social optimism scale and also for each subscale. In order to reduce the complexity, recently a revised scale composed of bipolar items was presented (Schweizer & Schreiner, in press). In constructing this scale pairs of items were merged in order to avoid the effect of item wording. In this process the number of items was reduced from 24 to 12. Because of the hierarchical structure the items of the bipolar measure of social optimism could be assumed to be semi-homogeneous.

This scale was applied to an internet sample of 769 participants. The mean age of the sample was 24.6 years of age (SD=4.59). It included 572 females and 197 males. Since mailing lists of various psychological institutes were used for addressing people, it can be assumed that the majority of participants were psychology students. An e-mail with an invitation to participate in the study was sent to the members of these lists. Only complete data sets were included in the investigations reported in the following paragraphs.

In the first step the data were investigated by conventional methods of item analysis. This investigation required the computation of part-whole correlations and Cronbach's Alpha. In the second step there was confirmatory factor analysis according to the standard model, which means on the basis of the congeneric model assuming one dimension. Finally, the bifactor model was applied. In applying the bifactor model specific latent variables were considered additionally for representing optimism concerning economic affairs (5 items), optimism concerning predicted disasters (3 items) and optimism concerning the future controllability of violence (4 items).

## Conventional item analysis

The idea of the part-whole correlations is that there is one general source of responding to items. Therefore, part-whole correlations can provide some kind of comparison level for the loading on a general latent random variable. The part-whole correlations are presented in the first column of Table 1.

Apparently, the part-whole correlations vary between .37 and .62. Cronbach's Alpha for the 12 items was .84 which is quite good. Furthermore, the investigation of the contributions of the individual items showed that the removal of no one of the items would lead to an increase of Cronbach's Alpha. Consequently, according to the conventional methods of item analysis there was no real reason for removing one of the items.

**Table 1:**
Part-whole Correlations, Completely Standardized Factor Loadings According to Congeneric and c-bifactor Models for the 12 Social Optimism Items

| Item | Part-whole correlation | Loadings for congeneric model | Loadings for c-bifactor model |
|---|---|---|---|
| 1 | .47 | .54 | .49 |
| 2 | .50 | .56 | .54 |
| 3 | .38 | .45 | .42 |
| 4 | .45 | .52 | .43 |
| 5 | .50 | .60 | .43 |
| 6 | .56 | .65 | .69 |
| 7 | .54 | .64 | .49 |
| 8 | .63 | .74 | .64 |
| 9 | .37 | .41 | .43 |
| 10 | .62 | .71 | .76 |
| 11 | .59 | .66 | .65 |
| 12 | .50 | .56 | .61 |

[1] Only the loadings on the general social optimism latent variable are given.

## Confirmatory factor analysis according to the standard model

Confirmatory factor analysis was performed by means of the maximum likelihood estimation method in using LISREL (Jöreskog & Sörbom, 2001). The completely standardized loadings are given in the second column of Table 1. These loadings vary between .41 and .71. Each one of them reaches the level of significance. Although the results for the individual items are favorable, the model shows a degree of fit that is not really good:

$\chi^2(54)$= 1,041.60 ($p<.01$), RMSEA=.154, SRMR=.081, GFI=.96, CFI=.95, NNFI=.93 (for interpreting the fit statistics see Schweizer, 2010). Especially, the $\chi^2$ and REMSEA results are far out of the acceptable ranges. As a consequence, the model must be rejected. It does not represent the data well although the loadings show a reasonable size.

### Confirmatory factor analysis according to the c-bifactor model

In order to conduct confirmatory factor analysis according to the c-bifactor model, three specific latent variables were considered. These latent variables represented optimism concerning economic affairs, optimism concerning predicted disasters and optimism concerning the future controllability of violence and received loadings from the corresponding items. The third column of Table 1 includes the completely standardized loadings on the general social optimism latent variable. The sizes of these loadings vary between .42 and .76. Each one reaches the level of significance. Furthermore, the investigation of the bifactor model reveals a good model fit: $\chi^2(43)$= 123.73 ($p<.01$), RMSEA=.049, SRMR=.032, GFI=.99, CFI=1.00, NNFI=1.00. Only the $\chi^2$ falls in the range between acceptable and good results whereas all the other fit statistics are within the range of good results. However, the large $\chi^2$ is probably due to the large sample size since $\chi^2$ is sensitive to sample size.

Comparing the results achieved for the standard model and the c-bifactor model clearly indicates an improvement in model fit due to the consideration of additional latent variables according to the lower-level structure. This improvement is also obvious from comparing the AIC statistics. The AIC of the standard model is 1,089.60 and of the c-bifactor model 193.73. There is a really considerable improvement.

## Discussion

After over 100 years of psychological research in some subject areas a quite advanced stage has been reached that is characterized by complex structures, as for example the subject areas of ability and personality. These structures integrate various constructs and specify the relationships among these constructs. Many structures are hierarchical structures including several levels. The constructs assigned to these levels show various degrees of generality and, as a consequence, can be assumed to be characterized by different degrees of homogeneity. Lower-level constructs can be assumed to show higher degrees of homogeneity than upper-level constructs. The constructs of the lowest level should show the highest degree of homogeneity.

The hierarchical structure can cause a problem in test construction if the methods for constructing measures concentrate on the maximization of homogeneity. Such methods preferably lead to favorable results for measures that represent lower-level constructs. In contrast, in measures representing higher-level constructs the probability of failure is high. Actually, success can only be expected in the case that such a measure is composed of items referring to different lower-level constructs. If there is no overlap between the items, it should also be possible to end up with a homogeneous measure. However, this possibility is quite restricted since in assuming that there should be at least 10 to 12 items a corresponding number of lower-level constructs must be available. Therefore, this situation is not really a realistic possibility.

The difference that is obvious from the description of the consequence of having hierarchies led to the proposal of the distinction of homogeneous and semi-homogeneous measures. Homogeneous measures are basically one-dimensional. Such measures mainly represent lower-level constructs. In contrast, semi-homogenous measures are characterized by one general and dominating dimension besides indications of other dimensions. Furthermore, there is the restriction that the other dimensions may only apply to subsets of the items. However, the dominating dimension must clearly be superior to the other dimensions.

The methods available for the construction of measures have different properties and, as a consequence, are appropriate for generating either homogeneous or semi-homogeneous measures. The theoretical reasoning showed that confirmatory factor analysis according to the congeneric model of measurement is most appropriate for the construction of homogeneous measures. This is no surprise since the congeneric model of measurement clearly states that there should be one latent variable only. For constructing a semi-homogeneous measure a basic modification of this model is necessary. This modification transforms the congeneric model into a c-bifactor model. In applied research there are already some examples of corresponding modifications of the congeneric model (e. g., Chen, West, & Sousa, 2006; Gignac, 2008; Schweizer & Rauch, 2008). The comparison of results achieved by means of the two models even enables the investigation of the question whether a given measure is either homogeneous or truly semi-homogeneous. In contrast, the conventional methods of test construction do not allow the distinction of the two types of homogeneity.

The results of investigating social optimism data confirm the results of comparing the models and methods according to their theoretical properties. Social optimism shows a known hierarchical structure. The congeneric model ignores this structure and assumed that the data can be represented by one latent variable. This model failed. The transformation into a c-bifactor model provided the opportunity to consider the lower-level structure. As expected, this model was successful. It showed a good model fit and yielded loading with an acceptable size for the general latent variable. Furthermore, the comparison of the models made obvious that the measure is actually semi-homogeneous. The results for the conventional models indicated a good quality of the measure.

It remains to consider the consequences for assessment. If the loadings on the additional latent variables are minor, there is a situation comparable to the situation after having conducted exploratory factor analysis. In exploratory factor analysis the items load on all the factors, and there is no problem as long as the pattern of loadings shows simple structure. Therefore, if the loadings on the general latent variable surmount the other loadings, the usual scoring function may be applied. In another case it would be advisible to select the factor scores for scoring.

## References

Bock, R. d., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: application of the EM algorithm. *Psychometrika, 46*, 443-459.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.

Carmines, E. G., & McIver, J. P. (1981). Analyzing models with unobservable variables: Analysis of covariance structures. In G. W. Bohrnstedt & E. F. Borgatta (Eds.), *Social measurement* (pp. 65-115). Berverly Hills, CA: Sage.

Carroll, J. B. (1993). *Human cognitive abilities*. New York: Cambridge University Press.

Chen, F. F., West, S. G., & Sousa, K. H. (2006). A comparison of bifactor and second-order models of quality of life. *Multivariate Behavioral Research, 41*, 189-225.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*, 297-334.

DiStefano, C., & Motl, R. W. (2006). Further investigating method effects associated with negatively worded items on self-report surveys. *Structural Equation Modeling, 13*, 440-464.

Eysenck, H. J. (1967). *The biological basis of personality*. Springfield, IL: Charles C. Thomas.

Ferrando, P. J. (2005). Factor analytic procedures for assessing social desirability in binary items. *Multivariate Behavioral Research, 40*, 331-349.

Ferrando, P. J., & Lorenzo-Seva, U. (2005). IRT-related factor analytic procedures for testing the equivalence of paper-and-pencil and internet-administered questionnaires. *Psychological Methods, 10*, 193-205.

Gibbons, D. G., Bock, D., Hedeker, D., Weiss, D. J., Segawa, E., Bhaumik, D. K., Kupfer, D. J., Frank, E., Grochocinski, V. J., & Stover, A. (2007). Full-information item bifactor analysis of graded response data. *Applied Psychological Measurement, 31*, 4-19.

Gignac, G. E. (2008). Higher-order models versus direct hierarchical models: g as superordinate or breadth factor? *Psychology Science Quarterly, 50*, 21-43.

Gulliksen, H. (1959). Theory of mental tests. New York: Wiley & Sons.

Hartig, J., Hölzel, B., & Moosbrugger, H. (2007). A confirmatory analysis of item reliability trends (CAIRT): differentiating true score and error variance in the analysis of item context effects. *Multivariate Behavioral Research, 42*, 157-183.

Holzinger, K. J., & Swineford, F. (1937). The bi-factor method. *Psychometrika, 2*, 41-54.

Jöreskog, K. G. (1971). Statistical analysis of sets of congeneric tests. *Psychometrika, 36*, 109-133.

Jöreskog, K., & Sörbom, D. (2001). *LISREL 8.5: User's reference guide*. Chicago, IL: Scientific Software International.

Kenny, D. A., & Kashy, D. A. (1992). Analysis of the multitrait-multimethod matrix by confirmatory factor analysis. *Psychological Bulletin, 112*, 165-172.

Kline, R. B. (2005). *Principles and practices of structural equation modeling* (pp. 133-145). 2nd ed. New York: The Guilford Press.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental scores*. Reading, Mass.: Addison-Wesley.

Lucke, J. F., (2005). The alpha and the omega of congeneric test theory: an extension of reliability and internal consistency of heterogeneous tests. *Applied Psychological Measurement, 29*, 65-81.

Marsh, H. W., & Bailey, M. (1991). Confirmatory factor analyses of multitrait-multimethod data: A comparison of alternative models. *Applied Psychological Measurement, 15*, 47-70.

McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, NJ: Lawrence Erlbaum Associates.

Rauch, W., Schweizer, K., & Moosbrugger, H. (2007). Method effects due to social desirability as a parsimonious explanation of the deviation from unidimensionality in LOT-R scores. *Personality and Individual Differences, 42*, 1597-1607.

Rauch, W., Schweizer, K., & Moosbrugger, H. (2008). An IRT analysis of the personal optimism scale. *European Journal of Psychological Assessment, 24*, 49-56.

Raykov, T. (1997). Scale reliability, Cronbach's coefficient alpha, and violations of essential tau-equivalence with congeneric components. *Multivariate Behavioral Research, 32*, 329-353.

Raykov, T. (2001). Estimation of congeneric scale reliability using covariance structure analysis with nonlinear constraints. *British Journal of Mathematical and Statistical Psychology, 54*, 69-76.

Schweizer, K. (2010). Guidelines concerning the modeling of traits and abilities in test construction. *European Journal of Psychological Assessment, 26*, 1-2.

Schweizer, K., Moosbrugger, H., & Schermelleh-Engel, K. (2003). Models for investigating hierarchical structures in differential psychology. *Methods of Psychological Research Online, 8*, 159-180.

Schweizer, K. & Rauch, W. (2008). An Investigation of the Structure of the Social Optimism Scale in Considering the Dimensionality Problem. *Journal of Individual Differences, 29*, 223-230.

Schweizer, K., & Schreiner, M. (2010). Avoiding the effect of item wording by means of bipolar instead of unipolar items: an application to social optimism. *European Journal of Personality*.

Schweizer, K., Schreiner, M., & Gold, A. (2009). The confirmatory investigation of APM items with loadings as a function of the position and easiness of items: a two-dimensional model of APM. *Psychology Science Quarterly, 51*, 47-64.

Steyer, R., Schmitt, M. J., & Eid, M., (1999). Latent trait-state theory and research in personality and individual differences. *European Journal of Personality, 13*, 389-408.

Vautier, S., & Raufaste, E., & Cariou, M. (2003). Dimensionality of the Revised Life Orientation Test and the status of filler items. *International Journal of Psychology, 38*, 390-400.