Review of Psychology,
2011, Vol. 18, No. 1, 3-11 UDC 159.9

The hypothesis-based investigation of patterns of relatedness by means of confirmatory factor models: The treatment levels of the Exchange Test as example

MICHAEL SCHREINER and KARL SCHWEIZER

This paper presents a method for formalizing hypotheses on the outcomes of performance according to different treatment levels that can be investigated by means of confirmatory factor models. This method includes several steps starting from the general research hypotheses and the characteristics of the stimulated processes and ends up with a pattern of relatedness that can be integrated into a confirmatory model for an investigation. It is a fixed-links model adapted to the pattern of relatedness with the factor loadings constrained according to this pattern. This method is demonstrated in a measure of working memory capacity that includes several treatment levels. Three alternative hypotheses and the corresponding patterns of relatedness are considered for an investigation. The data for the investigation originate from an internet study. The results demonstrate the usefulness of this systematic approach to the investigation of assumed patterns. It turns out that in the internet data one pattern is clearly superior to the other patterns.

Key words: confirmatory factor analysis, fixed-links model, working memory, internet data

Confirmatory factor analysis was originally proposed as a method for the confirmatory investigation of the structure of covariance matrices (Jöreskog, 1970). In such investigations the focus was on the correct assignment of manifest to latent variables, and it has proven to be a valuable research tool in a large number of applications. The theoretical importance of this focus results from the association of the latent variable with a specific concept that is defined in terms of a psychological construct (MacCorquodale & Meehl, 1948). Great advances in psychological science have been achieved by means of this research tool.

However, confirmatory factor analysis is not really restricted to the standard approach that means investigating the correct assignment of manifest to latent variables. Starting with the work by McArdle (1986, 1988; McArdle & Epstein, 1987) and Meredith and Tisak (1984, 1990) confirmatory factor analysis has been refined by fixing loadings to specific sizes for investigating the presence of specific relationships mostly described as curves in repeated-measures data. This refinement has led to so-called growth curve and latent curve models (Duncan & Duncan, 2004; McArdle, 2009). These models proved to be very useful in the area of developmental and educational research. Furthermore, there are fixed-links models (Schweizer, 2006a, 2006b, 2007, 2008, 2009) that aim at the hypothesis-guided decomposition of variances and covariances. Within this approach hypotheses can give rise to various types of specific patterns of the relatedness of the manifest and latent variables that provide the outset for such an investigation. Growth curve and latent curve models as well as fixed-links models share the same mathematical foundations although they have been developed in different directions.

An example that is especially appropriate for highlighting the special characteristics of such patterns shall provide the outset for the presentation of the rationale. The example is a measure of working memory capacity (Schweizer, 1996) that includes several treatment levels, which give rise to precise relational expectations based on cognitive theory, which can easily be summarized by a formal hypothesis. This measure requires the mental exchange of simple figures to bring two series of simple figures to a match. Figure 1 gives an example of such a pair of two series.

Michael Schreiner, Department of Psychology, Goethe University Frankfurt, Mertonstr. 17, 60056 Frankfurt a. M., Germany.
E-mail: Schreiner@psych.uni-frankfurt.de (the address for correspondence);

Karl Schweizer, Goethe University Frankfurt, Frankfurt a. M., Germany.

Acknowledgement
Parts of the research were supported by Deutsche Forschungsgemeinschaft, Kennedyallee 40, 53175 Bonn, Germany (SCHW 402/14-1)

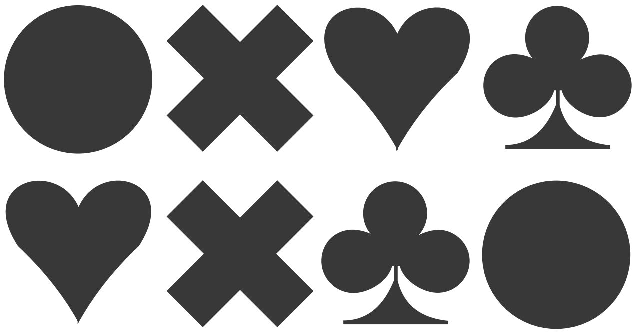

Figure 1. Sample stimuli of a trail of the Exchange Test (fifth treatment level demanding four exchanges).

The demands of five treatment levels interpreted in the light of working memory theory (Baddeley, 1986) provide the basis of the expectations. Each treatment level is characterized by the necessity to perform a specific number of mental exchanges of simple figures in combination with related storage operations. Different accuracy levels in responding to the stimuli of different treatment levels are suggested by available models of working memory. Consequently, the accuracy scores associated with a treatment level can be perceived as the result of a function *f* with the treatment level as argument such that

Accuracy score (Level 1) = f [Treatment (Level 1)]
Accuracy score (Level 2) = f [Treatment (Level 2)]
Accuracy score (Level 3) = f [Treatment (Level 3)]
Accuracy score (Level 4) = f [Treatment (Level 4)]
Accuracy score (Level 5) = f [Treatment (Level 5)].

The function is assumed to reflect the cognitive processes according to theory of working memory. Confirmatory factor analysis of the accuracy scores achieved by the cognitive measure according to the standard approach would require examining whether the same latent variable is underlying the performance in completing the trials of the five treatment levels. Although the factor loadings might indicate different degrees of association between the accuracy scores and the latent variable, a real investigation of precise relational expectations would not be possible this way.

For an appropriate investigation of the expectations by means of refined confirmatory factor analysis it is necessary to transform these expectations into a formal representation that is a pattern of relatedness and can be integrated into a confirmatory factor model. For this purpose there must be precise relational expectations, i.e. numbers that establish specific relationships between the treatment levels. This paper elaborates on this approach, and its usefulness is demonstrated in data achieved by means of the mentioned measure of working memory.

Hypotheses as patterns of assignments

The confirmatory nature of the standard approach of confirmatory factor analysis is dependent on the availability of a hypothesis concerning the underlying structure. In the simplest case the hypothesis suggests one latent variable as systematic source, which mainly accounts for the variances and covariances of a set of manifest variables. Taking the perspective of the individual manifest variable the hypothesis suggests that there is a considerable relationship of manifest and latent variables. The elaboration of this perspective gave rise to the congeneric model of measurement (Jöreskog, 1971). It is this model of measurement that is usually selected for the purpose of test construction in confirmatory factor analysis. This model of measurement has even given rise to a specific test theory that is based on the true-score model: the congeneric test theory (Jöreskog, 1971; Lucke, 2005; McDonald, 1999; Raykov, 1997, 2001).

Because of the great importance of this model of measurement a formal description needs to be provided. Assume that $\mathbf{Y}$ is the $P \times 1$ vector of observations, μ the $P \times 1$ vector of true item-specific components, $\mathbf{\Lambda}$ the $P \times Q$ matrix of loadings, $\boldsymbol{\eta}$ is the $Q \times 1$ vector of true person-specific components and $\boldsymbol{\varepsilon}$ the $P \times 1$ vector of error components. Then the congeneric model of measurement is given by

$$\mathbf{Y} = \boldsymbol{\mu} + \mathbf{\Lambda}\boldsymbol{\eta} + \boldsymbol{\varepsilon}. \qquad (1)$$

Since in confirmatory factor analysis variances and covariances are investigated instead of raw scores, the true item-specific component is usually omitted, and the model of measurement is usually presented as model of $\mathbf{y}$ that is the $P \times 1$ vector of mean-standardized observations (Bollen, 1989, p.18):

$$\mathbf{y} = \mathbf{\Lambda}\boldsymbol{\eta} + \boldsymbol{\varepsilon}. \qquad (2)$$

Normally only this general version of the model of measurement is presented in order to give a description of the general characteristics of the investigation. However, if there are P items and one latent source, there are actually P specific models of measurement, and it is therefore reasonable to list most or all of these models of measurement in such a systematic treatment of the issue:

$$\begin{array}{ccccc} y_1 & = & \lambda_1\eta & + & \varepsilon_1 \\ y_2 & = & \lambda_2\eta & + & \varepsilon_2 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ y_p & = & \lambda_p\eta & + & \varepsilon_p \end{array}$$

Although these specific versions of the model of measurement combine a number of assumptions, there is a core that is considered as the hypothesis on which the investigation concentrates: the true component symbolized by η represents the only systematic source of responding to the items, and the fact that an item is related to this source is indicated by a substantial factor loading. This hypothesis is represented by the $P\times1$ vector **h**:

$$\mathbf{h} = \begin{bmatrix} h_1 \,(= \lambda_1) \\ h_2 \,(= \lambda_2) \\ . \\ . \\ h_p \,(= \lambda_p) \end{bmatrix}, \qquad (3)$$

where h_p ($p = 1,\ldots, P$) is the hypothesis-based constraint suggested for l_p ($p = 1,\ldots, P$) that are the factor loadings of the pth item on the latent variable. Usually the model is specified by estimating the loadings and error components and by fixing the variance of the latent variable. An alternative way is to fix one loading in combination with the estimation of the other loadings and the variance. However, this alternative shows a considerable dependency on the selection of the loading to be fixed (Schweizer, 2010a).

Hypotheses as patterns of relatedness

In some areas of the psychological science the expectations concerning the relationships of items and latent variables are rather precise, and sets of related expectations give rise to *patterns of relatedness*. Such expectations eliminate the vagueness that characterizes the relationship of latent variable and items in the standard approach of confirmatory factor analysis. Because of the precise relational expectations it is no longer necessary to search for an estimate of the relationship. Instead there is a proposal suggesting specific degrees of association, and the emphasis is on the evaluation of the appropriateness of this proposal. As a consequence, the estimation of factor loadings as major aim is replaced by the estimation of the variances of the latent variables (Schweizer, 2008).

The available patterns of relatedness enable the hypothesis-guided decomposition of the variances and covariances of the manifest variables. Such patterns can take the role of factor loadings, so that the constrained factor loadings guide the variance decomposition. Such loadings are addressed as constraints or fixed loadings. For investigating a pattern of relatedness it is insufficient to provide just the general model of measurement because the fixed loading of each specific model of measurement is an important part of the pattern of relatedness. So the size of the loading of each of the following specific models of measurement counts:

$$\begin{array}{ccccc} x_1 & = & \lambda_1\eta & + & \delta_1 \\ x_2 & = & \lambda_2\eta & + & \delta_2 \\ . & . & . & . & . \\ . & . & . & . & . \\ x_p & = & \lambda_p\eta & + & \delta_p \end{array}$$

The fixation is made obvious by printing the loading normal (i.e. not in italics). This set of specific models of measurement provides the basis of what is considered as the formal hypothesis. Again a $P\times1$ vector **h** which includes the constraints, that is, fixed loadings, as elements serves as the formal representation of the hypothesis:

$$\mathbf{h} = \begin{bmatrix} \mathrm{h}_1 \,(= \lambda_1) \\ \mathrm{h}_2 \,(= \lambda_2) \\ . \\ . \\ \mathrm{h}_p \,(= \lambda_p) \end{bmatrix}. \qquad (4)$$

The difference between the Equations 3 and 4 is illustrated by the font of the elements: normal (4) vs. italics (3). Since the elements of **h** are real numbers, there is no freedom to adapt to the specificities of the data. This means that the advantage of learning about the appropriateness of a very specific hypothesis is accompanied by the increased likelihood of failure since even a small deviation of the hypothesized pattern from the true pattern may lead to a failure.

Because of the constraint of the elements of **h** there are always a great number of alternative hypotheses. Assume $\mathbf{h}^*$ such that

$$\mathbf{h}^* = \begin{bmatrix} \mathrm{h}^*_1 \,(= \lambda^*_1) \\ \mathrm{h}^*_2 \,(= \lambda^*_2) \\ . \\ . \\ \mathrm{h}^*_p \,(= \lambda^*_p) \end{bmatrix}. \qquad (5)$$

Then $\mathbf{h}^*$ is an alternative of **h** if there is one pair of corresponding loading $\lambda_p\,(\in\mathbf{h})$ and $\lambda^*_p\,(\in\mathbf{h}^*)$ such that

$$\lambda_p \neq \lambda^*_p. \qquad (6)$$

The simultaneous test of several hypotheses

This section investigates the possibility of combining patterns of relationships associated with different hypotheses since single patterns may not be sufficient for achieving a good degree of model fit. Furthermore, there is a possibility that different processes that give rise to different patterns contribute to performance.

Because of the constraint of factor loadings in investigating patterns of relatedness there are favourable preconditions for the simultaneous investigation of several hypotheses. The model of the covariance matrix (Jöreskog, 1970) provides the outset for further considerations. Let $\boldsymbol{\Sigma}$ be the model of the $P \times P$ covariance matrix, $\boldsymbol{\Lambda}$ the $P \times Q$ matrix of factor loadings, $\boldsymbol{\Phi}$ the $Q \times Q$ covariance matrix of the latent variables, and $\boldsymbol{\Theta}$ the $P \times P$ diagonal matrix of error variances. Then the definition of the model of the covariance matrix is given by

$$\mathbf{\Sigma} = \mathbf{\Lambda\Phi\Lambda' + \Theta}. \quad (7)$$

This model was proposed for the case where each item loads on one latent variable only.

Fortunately, the loss of control over the latent variable is not a problem for models with fixed loadings. This fact becomes obvious by integrating two or more latent variables representing different hypotheses into one model. Assume Σ_1 and Σ_2 as two models of the $P \times P$ covariance matrix and two $P \times 1$ vectors λ_1 and λ_2 that represent different hypotheses. Furthermore, let ϕ_1 and ϕ_2 be the variances of two uncorrelated latent variables, and Θ_1 and Θ_2 the $P \times P$ diagonal matrices of error variances. Then the two hypotheses give rise to two specific models for the same covariance matrix:

$$\mathbf{\Sigma}_1 = \lambda_1\phi_1\lambda_1' + \Theta_1 \quad (8)$$

and

$$\mathbf{\Sigma}_2 = \lambda_2\phi_2\lambda_2' + \Theta_2. \quad (9)$$

Integrating the two hypotheses into one model of the covariance matrix means that the diagonal matrices of error variances Θ_1 and Θ_2 must be replaced by a new diagonal matrix of error variances Θ^*_{1+2}. Because of the independence assumption a rather simple way of integration is possible:

$$\mathbf{\Sigma}_{1+2} = \lambda_1\phi_1\lambda_1' + \lambda_2\phi_2\lambda_2' + \Theta^*_{1+2}. \quad (10)$$

It is apparent that the vectors representing the hypotheses can be retained without the slightest modification.

Equation 10 can easily be transformed into the standard notation of Equation 7. It is simply necessary to integrate loading vectors λ_1 and λ_2 into the matrix of factor loadings Λ:

$$\Lambda = [\lambda_1,\lambda_2], \quad (11)$$

and to integrate the variances of the latent variables into the 2×2 covariance matrix of latent variables Φ:

$$\Phi = \begin{bmatrix} \phi_1 & 0 \\ 0 & \phi_2 \end{bmatrix}. \quad (12)$$

The investigation and comparison of hypotheses

The investigation of hypotheses according to the Equations 2 and 4 requires the integration of the corresponding vectors into confirmatory factor models and the computation of the model-fit statistics. In the next step the results are evaluated by considering the various criteria proposed for this purpose (see Kline, 2005). Depending on the result of the evaluation the hypothesis is either rejected or retained. There are no special criteria for either patterns of assignments or patterns of relatedness.

Models representing alternative hypotheses need to be compared with each other if each one of them passes the evaluation. Alternative models are either nested or non-nested models. Unfortunately, hypotheses based on a pattern of relatedness normally give rise to models that are non-nested and show the same degrees of freedom so that only the methods for comparing non-nested models can be considered.

Well-known statistics that apply to non-nested models are AIC and BIC. These fit indices are based on the chi-squares and take the complexity of the model into consideration. A comparison of models on the basis of these indices is a simple comparison of results according to size where the lower size is the more favorable result. It is not possible to consider any additional type of confidence interval.

More recently research that explored the properties of available fit indices has provided new opportunities for non-nested comparisons that even allow the consideration some type of confidence limit. Particularly, there is the comparative fit index (CFI) by Bentler (1990). Based on a simulation study Cheung and Rensvold (2002) consider a CFI difference of 0.01 a substantial difference. The results of this study appear to be convincing although the range of considered structures is a bit limited.

AN EXAMPLE FROM COGNITIVE PSYCHOLOGY

A measure of working memory capacity serves as an example. This measure stimulates exchange operations that are associated with storage operations and is denoted Exchange Test (Schweizer, 1996). It already served as an example in the introductory section. In completing the individual trials of this computer-based measure a few simple figures (circle, heart, spade, cross, etc.) arranged as a series showing a specific sequence have to be exchanged mentally. The exchanges are restricted to adjacent figures so that intermediary configurations are achieved and need to be temporarily maintained. The participants have to perform exchanges of adjacent figures until two series of four simple figures show an identical order. Furthermore, the participants have to count the number of exchanges. After the completion of the task the response button has to be pressed. Pressing the response button causes the removal of the figures from the screen, and the participant is asked to enter the amount of performed exchanges.

The difficulty characterizing this measure is ascribed to the increasing load on working memory (see Carpenter, Just, & Shell, 1990). In the Exchange Test a high number of exchanges and associated storage operations means a high load on working memory that can be assumed to decrease the accuracy in responding. The Exchange Test provides two types of performance measures, accuracy and reaction time. Since single exchange operations are very easy, storage problems are to be considered as the major source of errors.

The Exchange Test was revised for this study. It was transformed from a lab version into an internet version. Transformation means that it was adapted to the habits and customs of internet users. It includes the elimination of the most difficult treatment level and special feedback so that an experimenter is no longer necessary. We considered one modification as especially important: lowering the degree of difficulty. Therefore the most demanding treatment level was removed from the test. The original version included six different treatment levels whereas the internet version consists of only five treatment levels. The first treatment level requires one exchange per trial, the second treatment level two exchanges, the third and fourth treatment levels three exchanges and the fifth treatment level four exchanges. In order to have a monotonic increase of the number of exchanges, one of the two treatment levels requiring the same number of exchanges is usually considered as a distraction level that increases uncertainty in responding.

The hypotheses

It proved to be useful to consider two groups of processes (Schweizer, 2008). First, there are the *main processes* (MP) which are in the focus of the experimental treatment (e.g. search processes, memory processes, transformation processes). The rationale of the experiment suggests that the systematic modification of the treatment means a systematic change of the stimulation of the main processes, that is, repeated stimulation in the case of the Exchange Test. Second, there are the *subsidiary processes* (SP) which are not in the focus of interest and normally contribute to information processing only once (e.g. reception of starting signal, responding after the termination of main processing). These processes are necessary for completing the task but are not modified in combination with the treatment level.

The major difference between the two types of processes is that one should be influenced by the experimental treatment whereas the other is indifferent with respect to the experimental treatment. While it is possible to ignore the subsidiary processes in the framework of the experimental approach, they cannot be neglected in the differential approach since they are a source of systematic variance.

In the case of the Exchange Test the main processes are exchange and storage processes. An exchange is defined as the mental reordering of a pair of simple figures that are arranged adjacent to each other. Furthermore, the new arrangement of simple figures must be temporarily stored to be available for the further processing. The subsidiary processes stimulated by the Exchange Test are perceptual processes necessary for the uptake of information, a decision process determining the type of reaction and a motor process since a response button has to be pressed.

The main hypothesis associated with the Exchange Test is that the increase in the number of exchanges leads to an increasing load on the working memory and should therefore have an influence on the accuracy in task completion. However, the increase is restricted to the main processes so that it is necessary to consider two different types of true-scores: true-scores representing exchange processes and related storage processes and true-scores representing subsidiary processes. The set of treatment levels including the distraction level gives rise to the following set of specific models of measurement:

$$
\begin{aligned}
y_{\text{accuracy\_level\_1}} &= \tau_{\text{Exchange\_process\_1}} + \tau_{\text{subsidiary\_process\_1}} + \varepsilon_1 \\
y_{\text{accuracy\_level\_2}} &= \tau_{\text{Exchange\_process\_2}} + \tau_{\text{subsidiary\_process\_2}} + \varepsilon_2 \\
y_{\text{accuracy\_level\_3}} &= \tau_{\text{Exchange\_process\_3}} + \tau_{\text{subsidiary\_process\_3}} + \varepsilon_3 \\
y_{\text{accuracy\_level\_4}} &= \tau_{\text{Exchange\_process\_4}} + \tau_{\text{subsidiary\_process\_4}} + \varepsilon_4 \\
y_{\text{accuracy\_level\_5}} &= \tau_{\text{Exchange\_process\_5}} + \tau_{\text{subsidiary\_process\_5}} + \varepsilon_5
\end{aligned}
$$

Exchange-number hypothesis (A). In order to arrive at numbers for an investigation, it is necessary to relate the *t*s of the various levels to each other. There is a basic exchange process associated with a storage process that can be assumed to be called up a different numbers of times. In contrast, the subsidiary process is always the same:

$$
\begin{aligned}
y_{\text{accuracy\_level\_1}} &= 1\times \tau_{\text{Exchange\_process}} + 1\times \tau_{\text{subsidiary\_process}} + \varepsilon_1 \\
y_{\text{accuracy\_level\_2}} &= 2\times \tau_{\text{Exchange\_process}} + 1\times \tau_{\text{subsidiary\_process}} + \varepsilon_2 \\
y_{\text{accuracy\_level\_3}} &= 3\times \tau_{\text{Exchange\_process}} + 1\times \tau_{\text{subsidiary\_process}} + \varepsilon_3 \\
y_{\text{accuracy\_level\_4}} &= 3\times \tau_{\text{Exchange\_process}} + 1\times \tau_{\text{subsidiary\_process}} + \varepsilon_4 \\
y_{\text{accuracy\_level\_5}} &= 4\times \tau_{\text{Exchange\_process}} + 1\times \tau_{\text{subsidiary\_process}} + \varepsilon_5
\end{aligned}
$$

The numbers of this set of specific models of measurement give the formal hypothesis $\mathbf{h_A}$:

$$
\mathbf{h_A} = \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \\ 3 & 1 \\ 4 & 1 \end{bmatrix}. \tag{13}
$$

The hypothesis represented by Equation 13 is denoted exchange-number hypotheses.

Furthermore, it is necessary to specify Φ appropriately since the theoretical model is very clear concerning the relationship between the two groups of processes. They must be independent of each other.

$$
\Phi = \begin{bmatrix} \phi_{11} & 0 \\ 0 & \phi_{22} \end{bmatrix}.
$$

Equation 13 is the theory-based hypothesis that provides the outset for further reasoning.

Set-size hypotheses (B). Furthermore, there is a possibility that not the number of exchanges is important but the number of simple figures that need to be manipulated mentally. The reason is that the treatment levels also differ according to the number of simple figures that need to be exchanged. This number is addressed as set-size. Although this hypothesis is not based on a theory of working memory, there is common sense suggesting that the different sizes may count. Therefore, the following alternative $\mathbf{h}_B$ needs to be considered additionally:

$$\mathbf{h}_B = \begin{bmatrix} 2 & 1 \\ 3 & 1 \\ 3 & 1 \\ 4 & 1 \\ 4 & 1 \end{bmatrix}. \quad (14)$$

The hypothesis represented by Equation 14 is denoted set-size hypothesis. In the easiest treatment level only two simple figures need to be considered, whereas in the most difficult one all four figures need to be taken into account.

Probability-adjusted exchange-number hypothesis (C) (Schweizer, 2011). However, the hypothesis of Equation 13 may be unrealistic because in the first treatment level the participants may notice that there is no alternative and, therefore, may not perform the expected exchange. After completing a few trials the participants may conclude that it is sufficient to verify that there are two simple figures, which need to be exchanged, and to respond accordingly. This means that it is reasonable to distinguish between treatment levels associated with certainty and uncertainty concerning the number of exchanges. The adaptation of **h** leads to

$$\mathbf{h}_C = \begin{bmatrix} 0 & 1 \\ 2 & 1 \\ 3 & 1 \\ 3 & 1 \\ 4 & 1 \end{bmatrix}. \quad (15)$$

The hypothesis represented by Equation 15 is denoted probability-adjusted exchange-number hypothesis. All in all, there are three hypotheses, which are reasonable with respect to some arguments and may therefore be compared with each other.

The models

The confirmatory factor models included five manifest variables corresponding to the treatment levels of the Exchange Test and two latent variables. One latent variable represented the main processes and the other one the subsidiary processes. These latent variables were not allowed to correlate with each other since they were expected to account for different parts of the variances and covariances. In contrast to the loadings, the variances of these latent variables were set free for estimation.

The various models that were investigated differed according to the constraints that were included. These constraints are given by the Equations 13 to 15.

Furthermore, it turned out that the variance of the first treatment level was very small and, as a consequence, it

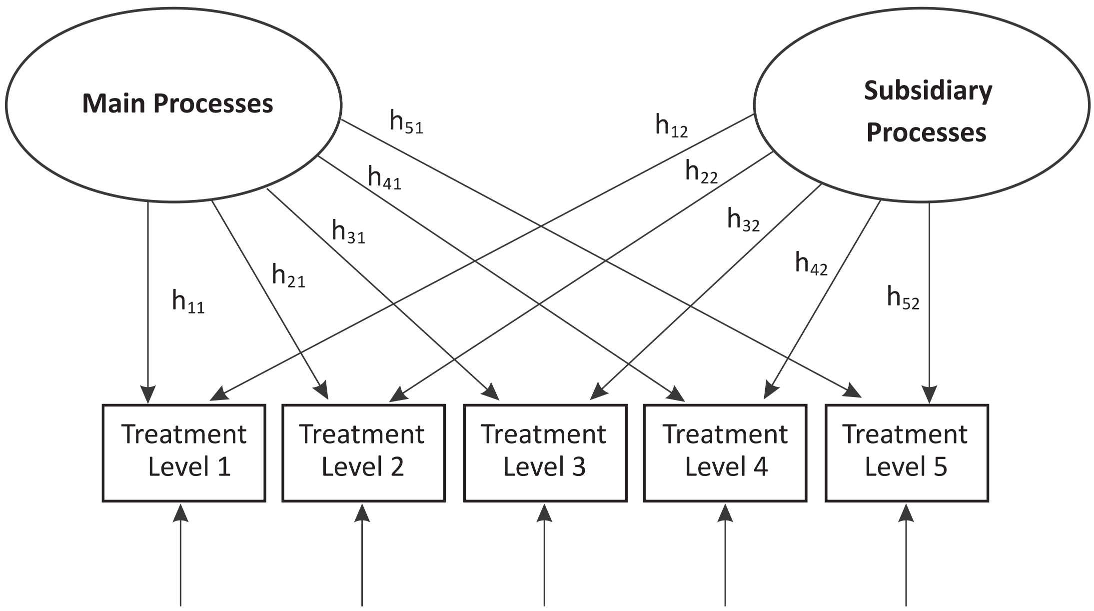

Figure 2. Confirmatory factor model with two latent variables representing the main and the subsidiary processes.

tended to be associated with a very small or unacceptable estimate of the error variance. Consequently, setting it to 1 would be inappropriate for the first treatment level, because participants would simply know the correct answer without actually performing mental exchanges. Therefore, the loading on the second latent variable, which does not reflect working memory capacity, was set to 0.01 instead of 1. Since there was no way of determining the exact amount, the new number was in a way arbitrary.

Additionally, a confirmatory factor model including the congeneric model of measurement was considered in order to also have an investigation concerning the pattern of assignments. In this case there was only one latent variable. The factor loadings were set free for estimation whereas the variance of the latent variable was set to one.

Data description

The psychology students of several universities were asked to participate in this study. The sample includes the data of 297 university students who were addressed via internet and agreed to participate. There were 98 males and 179 females. The mean age was 26.05 (SD = 5.88). The precondition for getting into the sample was producing a minimum of 50% correct responses for each treatment level.

The mean accuracy scores were 11.92, 11.19, 10.76, 10.58 and 8.84 for the first to fifth treatment levels in corresponding order. The standard deviations were 0.30, 1.05, 1.35, 1.54 and 1.74 in corresponding order.

The fit statistics for the models representing the hypotheses

The models with constraints and the standard model were investigated by means of LISREL (Jöreskog & Sörbom, 2001). The fit results (χ^2, χ^2/df, RMSEA, GFI, CFI, NNFI and AIC) for the models are presented in Table 1.

At first the evaluation concentrates on the two fit statistics that are considered the main statistics concerning model fit: the normed chi-squares (χ^2/df) and RMSEA. Normed chi-squares below 2 indicate a good model fit and below 3 an acceptable model fit, and RMSEAs equal to or below .05 are considered as good and equal to or below .08 as acceptable (see Schweizer, 2010b). According to these main statistics the set-size model and the congeneric model are not acceptable. Only the models representing the exchange-number hypothesis and the probability-adjusted exchange-number hypothesis remain. Each one of them shows an acceptable model fit according to the main fit statistics but not a good one. A comparison of the fit results of these models makes it obvious that the model according to the probability-adjusted exchange-number hypothesis does considerably better than the model representing the exchange-number hypothesis. Since in comparing non-nested models AIC and CFI should be consulted, the corresponding results need to be considered carefully. According to AIC the model of the probability-adjusted exchange-number hypothesis outperforms the other model since the results are 31.09 and 37.12 in corresponding order. Furthermore, the CFI difference is .05 which is considerably larger than .01. Obviously, the model according to the probability-adjusted exchange-number hypothesis provides the best account of the data.

Since the Exchange Test is considered as a measure of working memory capacity in the first place, the emphasis is on the representation of the exchange processes and the related storage processes. In contrast, the subsidiary processes are not of interest. Therefore, it was reasonable to try to find out how the model performs without the latent variable representing the subsidiary processes. Consequently, one latent variable was removed together with the corresponding column of the matrix of factor loadings. Furthermore, since in the model according to the probability-adjusted exchange-number hypothesis the accuracy of the first treatment level does not load on the latent variable representing the exchange processes and the related storage processes, the corresponding loading was simply set free for estimation. This model also showed an acceptable model fit (χ^2/dfs = 2.23, RMSEA = .065, GFI = .97, CFI = .92, NNFI = .90, AIC = 31.89). This is a very important finding since it clearly shows that the exchange processes and related storage pro-

Table 1
The fit statistics of the models including the various hypotheses concerning performance in the Exchange Test (N = 297)

Hypothesis characterizing the model	$\chi2$	df	$\chi2/df$	RMSEA	GFI	CFI	NNFI	AIC
Exchange-number hypothesis	23.12	8	2.89	.080	.97	.88	.86	37.12
Set-size hypothesis	25.75	8	3.21	.087	.97	.87	.84	39.75
Probability-adjusted exchange-number hypothesis	17.10	8	2.13	.062	.98	.93	.91	31.09
Congeneric hypothesis	15.07	5	3.01	.082	.98	.92	.85	35.07

Note. RMSEA = Root Mean Square Error of Approximation; GFI = Goodness-of-Fit Index; CFI = Comparative Fit Index; NNFI = Nonnormed Fit Index; AIC = Akaike Information Criterion.

cesses are indeed the most important systematic source of performance in completing the Exchange Test.

CONCLUSION

Two fundamentally different types of hypotheses can be investigated by using confirmatory factor analysis. These types are associated with different types of theories characterizing different stages of a theoretical development: theories that are very basic and aim at the identification of the basic concepts on the one hand, and theories suggesting a fine-grained structure of relationships among basic concepts on the other. In the former case the researchers are concerned with the simple question whether the observed phenomena represent the same latent attribute. In the latter case the researchers already have rather precise expectations concerning the relationships of the referents of the latent concepts. So the two types of hypotheses do not exclude each other but guarantee that confirmatory factor analysis contributes to the advancement of science at different levels of the theoretical development.

The systematic treatment of the approach associated with patterns of relatedness in this paper revealed a systematic way of formalizing hypotheses so that they can be investigated by means of confirmatory factor models. By applying it to the working memory task it became apparent that a set of models can be achieved that seem to show only minor differences: the model assuming exchange and related storage processes as a major source, the model assuming that the set-size counts, and the model assuming exchange and related storage processes as a major source with the exception of the first treatment level. However, despite the minor differences it was possible to clearly identify the best model. Obviously, the exchange and related storage processes are the major source of responding with the exception of the first treatment level once the participants realize that no actions needs to be taken.

The two approaches in confirmatory factor analysis aiming at either patterns of assignments or patterns of relatedness are closely associated with the treatment of factor loadings as estimates or constraints. The model of the covariance matrix provides the basis for both approaches. It is the most general model that can be further elaborated into different directions and reflects the foundations of factor analysis (Rao, 1958; Scher, Young, & Meredith, 1960; Tucker, 1958). The standard model that has dominated research for a long time elaborates in the direction of the patterns of assignments whereas the other model enables the investigation of patterns of relatedness.

REFERENCES

Baddeley, A.D. (1986). *Working memory*. Oxford: Clarendon Press.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*, 238-246.

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

Carpenter, P. A., Just, M. A., & Shell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices Test. *Psychological Review, 97*, 404-431.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, *9*, 233-255.

Duncan, T. E., & Duncan, S. C. (2004). An introduction to latent growth curve modeling. *Behavior Therapy*, *35*, 333-363.

Jöreskog, K. G. (1970). A general method for analysis of covariance structure. *Biometrika*, 57, 239–257.

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

Jöreskog, K. G., & Sörbom, D. (2001). *LISREL* (Version 8.5) [Computer Software]. Chicago: Scientific Software International.

Kline, R. B. (2005). *Principles and practices of structural equation modeling*. New York: The Guilford Press.

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.

McArdle, J. J. (1986). Latent variable growth within behavior genetic models. *Behavior Genetics*, *16*, 163-200.

McArdle, J. J. (1988). Growth curve analysis in contemporary psychological research. In J. R. Nesselroade & R. B. Cattell (Eds.), *Handbook of multivariate experimental psychology* (2nd ed., pp. 561-614). New York: Plenum.

McArdle, J. J. (2009). Latent variable modeling of differences and change with longitudinal data. *Annual Review of Psychology, 60*, 577-605.

McArdle, J. J., & Epstein, D. (1987). Latent growth curves within developmental structural equation models. *Child Development*, *58*, 110-133.

MacCorquodale, K., & Meehl, P. E. (1948). On a distinction between hypothetical constructs and intervening variables. *Psychological Review*, *55*, 95-107.

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

Meredith, W., & Tisak, J. (1984). *Tuckerizing curves*. Paper presented at the annual meeting of the Psychometric Society, Santa Barbara, CA.

Meredith, W., & Tisak, J. (1990). Latent curve analysis. *Psychometrika*, *55*, 107-122.

Rao, C. R. (1958). Some statistical methods for comparisons of growth curves. *Biometrics*, *14*, 1-17.

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.

Scher, A. M., Young, A. C., & Meredith, W. M. (1960). Factor analysis of the electrocardiograph. *Circulation Research*, *8*, 519-526.

Schweizer, K. (1996). The speed-accuracy transition due to task complexity. *Intelligence, 22,* 115-128.

Schweizer, K. (2006a). The fixed-links model in combination with the polynomial function as tool for investigating choice reaction time data. *Structural Equation Modeling*, *13*, 403-419.

Schweizer, K. (2006b). The fixed-links model for investigating the effects of general and specific sources. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences, 2,* 149-160.

Schweizer, K. (2007). Investigating the Relationship of Working Memory Tasks and Fluid Intelligence Tests by Means of the Fixed-Links Model in Considering the Impurity Problem. *Intelligence, 35*, 591-604.

Schweizer, K. (2008). Investigating experimental effects within the framework of structural equation modeling: an example with effects on both error scores and reaction times. *Structural Equation Modeling*, *15*, 327-345.

Schweizer, K. (2009). Fixed-links models for investigating experimental effects combined with processing strategies in repeated measures designs: a cognitive task as example. *British Journal of Mathematical and Statistical Psychology, 62*, 217-232.

Schweizer, K. (2010a). Improving the interpretability of the variances of latent variables by uniform and factor-specific standardization of loadings. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences, 6*, 152-159.

Schweizer, K. (2010b). Some Guidelines Concerning the Modelling of Traits and Abilities in Test Construction. *European Journal of Psychological Assessment, 26*, 1-2.

Schweizer, K. (2011). Probability-based and measurement-related hypotheses with full restriction for investigations by means of confirmatory factor analysis: an example from cognitive psychology. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*. doi:10.1027/1614-2241/a000033

Tucker, L. R. (1958). Determination of parameters of a functional relation by factor analysis. *Psychometrika*, *23*, 19-23.