Psychological Test and Assessment Modeling, Volume 56, 2014 (1), 45-59

On the ways of investigating the discriminant validity of a scale in giving special emphasis to estimation problems when investigating multitrait-multimethod matrices

Karl Schweizer[1]

Abstract

Discriminant validity is a valuable property of psychological scales that is usually investigated in the framework of the multitrait-multimethod approach. The establishment of discriminant validity demands the demonstration that the scale of interest representing a specific construct is unrelated to scales representing other constructs. The original implementation of the multitrait-multimethod approach demands a large number of comparisons among the correlations of a multitrait-multimethod design. More recently discriminant validity is investigated by means of confirmatory factor models including latent variables for the representation of constructs and methods. The process of arriving at a decision concerning discriminant validity in investigating multitrait-multimethod data is described. Downsizing the complexity of the model and the usage of a ridge option are proposed and applied for overcoming the estimation problems that frequently obstruct confirmatory factor analysis of multitrait-multimethod data.

Key words: disciminant validity, construct validity, construct, confirmatory factor analysis, multi-trait-multimethod matrix

[1] *Correspondence concerning this article should be addressed to:* Karl Schweizer, PhD, Department of Psychology, Goethe University Frankfurt, Grüneburgplatz 1, 60323 Frankfurt a. M., Germany; email: K.Schweizer@psych.uni-frankfurt.de

This paper presents descriptions of the major methods of establishing disrcriminant validity. It concentrates on the early correlational approach and the recent model-based approach that means confirmatory factor analysis of multitrait-multimethod data. Since the model-based approach of confirmatory factor analysis is plagued by estimation problems, special attention is given to the treatment of such problems. The method of downsizing the complexity of the model and the ridge option are considered for this purpose.

Discriminant validity is a property which psychological scales are expected to show. A scale owing this property has been proven not to correlate with scales representing constructs that are considered as unrelated to the construct that is represented by this scale. According to Campbell and Fiske (1959) discrimination is an important characteristic of an innovation in personality assessment since "[w]hen a dimension of personality is hypothesized, when a construct is proposed, the proponent invariably has in mind distinctions between the new dimensions and other constructs already in use" (p. 84). Furthermore, a distinction on the level of constructs demands for the check of the expected discrimination on the empirical level. Accordingly Campbell and Fiske state that "… distinctions are an important part of the validation process …" (p. 84).

However, discriminant validity is rarely considered in isolation and usually not classified as one of the major types of validity (Cizek, Rosenberg, & Koons, 2008). Mostly it appears in combination with convergent validity. The framework for this combination is construct validity. According to Cronbach and Meehl (1955) this concept was proposed as part of *technical recommendations* (1954) provided by an APA commission. Construct validity is closely linked to the concept of the nomological network specifying the relationships among a number of constructs (Ziegler, Booth, & Bensch, 2013). However, locating a construct within such a nomological network by means of investigations of discriminant validity is very trying since a large number of relationships have to be considered. In contrast, the investigation of convergent validity is much more focused than the investigation of discriminant validity, and as a consequence, convergent validity can be achieved in a more straight-forward way. So for quite a while and even nowadays (Schweizer, 2012), priority was given to the investigation of convergent validity.

It was the lack of trust in the outcome of investigating convergent validity that spurred on the interest in discriminant validity as part of the multitrait-multimethod approach proposed by Campbell and Fiske (1959). The multitrait-multimethod approach combines the investigation of the convergent validity of a scale with the investigation of the discriminant validity of the scale into one coherent research design. The introduction of the multitrait-multimethod approach was not only the consequence of considering distinctions but also of the insight that scales do not only reflect the constructs of interest but also the assessment method. The contribution of the assessment method to measurement even turned out to enlarge the correlations between scales in investigations of convergent validity if these scales were associated with the same assessment method. Given this situation the consideration of discriminant validity appeared to be essential for having a kind of comparison level in the evaluation of convergent validity.

After having introduced the multitrait-multimethod approach it appears necessary to point out an inconsistency that is a bit annoying at first view: the multitrait-multimethod approach serves the demonstration of construct validity and, therefore, the denotation as multiconstruct-multimethod approach would appear as the more appropriate one. But obviously Campbell and Fiske preferred to restrict the approach to traits, probably because they considered trait research as the major field of application of this approach. However, it turned out that this approach is also useful for other fields, as for example the fields of attitude research and of interest research. Therefore, in this paper the term construct is preferred to the term trait whenever possible.

Since its presentation the multitrait-multimethod approach has exerted a considerable influence on the evaluations of the validity of psychological scales. Concerning the statistical implementation of the multitrait-multimethod approach it is possible to distinguish two phases: the phase dominated by correlational analysis and the phase dominated by confirmatory factor analysis. There were also other implementations; but no one of the other implementations proved to be very influential. In the following sections the two major implementations of the approach are described in some detail.

The multitrait-multimethod approach as correlational analysis

The multitrait-multimethod approach is based on the expectation that the measurement of the construct of interest by means of a scale comprising a specific assessment method is valid if it is confirmed by different assessment methods and there is discrimination from the measurements of other constructs. Confirmation gives rise to convergent validity and discrimination to discriminant validity. Since assessment methods can be expected to perform in the same way in the investigations of convergent and discriminant validity, the comparison of the results achieved in investigating both types of validity can be expected to reveal the influence of the assessment method. So both confirmation and discrimination are indispensable components of the multitrait-multimethod approach.

A major characteristic of this approach is the systematic combination of constructs and assessment methods. Each construct which is considered needs to be operationalized in considering each one of the assessment methods selected for the study of the construct validity of the scale. Multitrait-multimethod designs for three constructs and three assessment methods are recommended. An example of a design following this recommendation was presented by Marsh (1989). It included the general school self-concept, the verbal self-concept and the math self-concept as constructs and three different questionnaires as assessment methods.

The original implementation of the multitrait-multimethod approach requires that data collected according to such a multitrait-multimethod design are used for the computation of correlations and for preparing a correlation matrix that shows a systematic structure so that it is possible to distinguish between different parts: triangles and blocks. Table 1 provides the structure of such a correlation matrix.

Table 1:
Multitrait-multimethod Matrix Including Three Constructs and Three Assessment Methods

		Method 1			Method 2			Method 3		
	Traits	A_1	B_1	C_1	A_2	B_2	C_2	A_3	B_3	C_3
	A_1	$(r_{A_1\tilde{A}_1})$								
Method 1	B_1	$r_{B_1A_1}$	$(r_{B_1\tilde{B}_1})$							
	C_1	$r_{C_1A_1}$	$r_{C_1B_1}$	$(r_{C_1\tilde{C}_1})$						
	A_2	$r_{A_2A_1}$	$r_{A_2B_1}$	$r_{A_2C_1}$	$(r_{A_2\tilde{A}_2})$					
Method 2	B_2	$r_{B_2A_1}$	$r_{B_2B_1}$	$r_{B_2C_1}$	$r_{B_2A_2}$	$(r_{B_2\tilde{B}_2})$				
	C_2	$r_{C_2A_1}$	$r_{C_2B_1}$	$r_{C_2A_1}$	$r_{C_2A_2}$	r_{C_2B2}	$(r_{C_2\tilde{C}_2})$			
	A_3	$r_{A_3A_1}$	$r_{A_3B_1}$	$r_{A_3C_1}$	$r_{A_3A_2}$	$r_{A_3B_2}$	$r_{A_3C_2}$	$(r_{A_3\tilde{A}_3})$		
Method 3	B_3	$r_{B_3A_1}$	$r_{B_3B_1}$	$r_{B_3C_1}$	$r_{B_3A_2}$	$r_{B_3B_2}$	$r_{B_3C_2}$	$r_{B_3A_3}$	$(r_{B_3\tilde{B}_3})$	
	C_3	$r_{C_3A_1}$	$r_{C_3B_1}$	$r_{C_3C_1}$	$r_{C_3A_2}$	$r_{C_3B_2}$	$r_{C_3A_2}$	$r_{C_3A_3}$	$r_{C_3B_3}$	$(r_{C_3\tilde{C}_3})$

Note. The numbers given in parentheses are reliability coefficients.

The considered scales are identified by a combination of a letter (A, B, C) that refers to an assessment method and a number (1, 2, 3) that refers to a construct. The triangles and blocks serve the identification of special areas of the matrix. These areas include correlations of the same type. The areas surrounded by solid lines include the correlations among scores obtained by the same assessment method. They are addressed as heterotrait-monomethod triangles. Furthermore, there are the areas surrounded by dashed lines. They include the correlations between different traits obtained by different assessment methods. They are addressed as heterotrait-heteromethod triangles. Moreover, there are correlations outside of these triangles and outside of the main diagonal. They are monotrait-heteromethod correlations that are arranged as validity diagonals. Finally, there are the reliabilities of the main diagonal. Campbell and Fiske distinguish between three parts of the main diagonal. Each part is associated with one assessment method and is addressed as reliability diagonal. The blocks are less prominent. There are monomethod and heteromethod blocks. Heteromethod blocks are composed of two heterotrait-heteromethod triangles and a validity diagonal.

A large number of comparisons can be conducted between the various correlations of the matrix in evaluating convergent and discriminant validity, and there are specific expectations. The reliability diagonals should include the largest correlations. Furthermore, it would be good if the correlations of the validity diagonals would surmount the remaining correlations. Since these expectations are a bit unrealistic for correlations of random variables, Campbell and Fiske present four criteria for the evaluation of a multitrait-multimethod matrix. The first criterion demands that the correlations of the validity diagonals are statistically significant and sufficiently large. According to the second criterion each correlation of a validity diagonal should be larger than any other correlation of the same row or column with the exception of the correlations of validity or reliability diagonals. The third criterion is concerning the comparisons of correlations of scales of the same trait obtained by different assessment methods with correlations of scales representing different traits and being due to the same assessment method. In such comparisons the correlation of scales of the same trait obtained by different assessment methods should always be the larger one. Finally, there is the fourth criterion demanding that each heterotrait triangle should show the same pattern of numbers. If a multitrait-multimethod matrix is in agreement with these guidelines, construct validity is considered as demonstrated that includes convergent and discriminant validity.

Despite the advantages of the multitrait-multimethod approach and the important role that is assigned to discriminant validity there is also a major weakness. The weakness is the dependency of the majority of constructs among each other. So there are only a few underlying personality dimensions that are independent of each other. A lot of work suggests the existence of only five basic dimensions (e.g., Norman & Goldberg, 1966; Tupes & Christal, 1961). Furthermore, frequently the considered constructs originate from the same hierarchical structure of personality so that only partial independence can be assumed. As a consequence, the sizes of the correlations computed for the evaluation of discriminant validity are frequently larger than zero and closer to the estimates of convergent validity than would be desirable.

Although the described procedure of evaluating a multitrait-multimethod matrix is very useful in making the influence of the assessment methods on the correlations between scales representing constructs obvious, it suffers from its high degree of complexity and the possibility of ambiguous outcomes. Therefore, it was characterized as an “informal” procedure (Cudeck, 1988). Another point is that the procedure can be expected to do quite well if all the reliabilities show more or less the same size whereas reliabilities of considerably differing sizes can be expected to do less well. Furthermore, small differences between correlations may reflect random error. The appropriate way would be to consider confidence intervals in comparing the correlations. Moreover, there is the disadvantage of having to compare a large number of correlations with each other in the evaluation of the multitrait-multimethod matrix (Schweizer, 2012). It means a distortion of the probability level. Some of the substantial results may simply be due to chance.

The multitrait-multimethod approach as confirmatory factor analysis

Confirmatory factor analysis is generally well suited for the investigation of the structure of a covariance or correlation matrix. However, the multitrait-multimethod approach poses special demands to confirmatory factor analysis that are incompatible with the standard model. Different types of models have been proposed for meeting these demands (Byrne & Goffin, 1993). The standard version of confirmatory factor analysis is the congeneric model of measurement (Jöreskog, 1971) that assumes one latent source of responding to the items of a scale. Since the multitrait-multimethod approach suggests the construct and assessment method as two independent sources of responding, in most attempts of adapting confirmatory factor analysis to the special demands of the multitrait-multimethod approach the congeneric model is replaced by a bifactor model (Chen, West, & Sousa, 2006; Reise, Morizot, & Hays, 2007; Schweizer, Altmeyer, Reiß, & Schreiner, 2010) that is characterized by manifest variables loading on two latent variables instead of on one latent variable only. The bifactor model also proved to be useful for the representation of effects that have been found to bias assessment, as for example item-position effect (Schweizer & Ren, 2013) and the wording effect (DiStefano & Motl, 2006).

Confirmatory factor analysis means a change of the focus of the investigation. Whereas in correlational analysis the focus is on the sizes of the correlations, in confirmatory factor analysis it is on the fit of the model to the data. The model is constructed according to the assumptions characterizing the multitrait-multimethod design. For example, there are sets of scales that are expected to represent the same constructs. Each of these constructs is assigned a latent variable that receives loadings from the scales assumed to represent this construct, and each method is associated with a latent variable with loadings from scales comprising the corresponding assessment method. A good model-data fit confirms these assumptions whereas a bad model-data fit calls them into question. In the case of a bad model-data fit there is at least one assumption that is not in agreement with the data. The specification of the model is called into question. There must be at least one misspecification.

The correct specification of the model is a crucial part of multitrait-multimethod modeling, and alternative specifications play a major role in the evaluation of a multitrait-multimethod matrix. Widaman (1985) proposed a general model characterized by correlated latent variables representing the constructs respectively the assessment methods as outset for the investigation of construct validity (CTCM). This model is only restricted as far as correlations between trait and method latent variables and among error components were not allowed. Figure 1 illustrates this model.

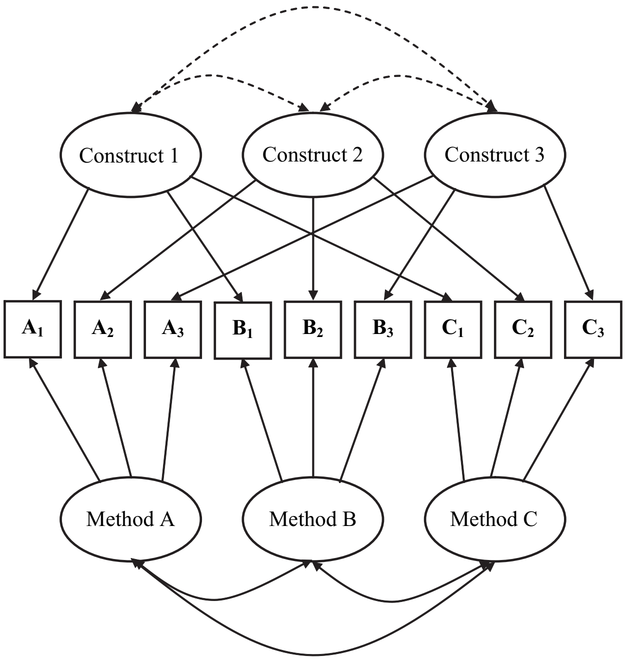

Figure 1:
General model representing the structure of a multitrait-multimethod design including three constructs and three traits.

Starting from this model it is possible to obtain various nested models by constraining model parameters to zero or one. For example, it may be suspected that two of the considered assessment methods do not really differ from each other so that contrasting these methods is not useful for the identification of method influence. Setting the correlation between the latent variables representing these assessment methods equal to one creates a model that is nested to the general model. A major advantage of nested models is that they can be compared by means of the chi-square difference test. An insignificant result indicates that both models account equally well for the data whereas in the other case a difference is signified. In the example an insignificant result confirms the supposition that the two methods do not differ from each other and signifies that the methods are useless for the identification of method variance.

The investigation of discriminant validity in considering the whole matrix. Since it is the aim to investigate discriminant validity that demands the enlargement of the research design by considering scales representing other constructs, the establishment of discriminant validity must concentrate on the discrimination regarding the major parts of the model. First, there is the discrimination between the latent variables representing different constructs. Discriminating between these constructs means distinction. According to Campbell and Fiske (1959) distinction is an important property of a scale that needs to be demonstrated, and the demonstration establishes discriminant validity. The demonstration requires the comparison of the general model with the model that assumes a general lack of discrimination. The assumption that there is no discrimination between the different constructs can be realized by setting the correlation between the corresponding latent variables equal to 1. This provision implicitly means that the scale of interest is indifferent with respect to the constructs. The resulting model is nested to the general model. Such nested models can be compared by means of the chi-square difference test. An insignificant result achieved in such a test indicates that both models account equally well for the data whereas in the other case of a significant result a difference between the models is signified. In the case of an insignificant result there is no discrimination between the constructs. The establishment of discriminant validity demands the observation of a substantial difference between the models.

The discrimination between the assessment methods is also important since it can make obvious whether the different assessment methods do what they are expected to do: the identification of method variance and the separation from construct variance. The logic of the investigation of the effects of the assessment methods is the same as with the constructs. Two models are considered. One model corresponds to the general model whereas the other model assumes that the latent variables associated with the assessment methods correlate perfectly. The chi-square difference test is employed for finding out whether these models differ substantially from each other. A substantial difference signifies that method variance is identified and does not impair the evaluation of discriminant validity. In contrast, in the case of an insignificant result there is the danger that the method latent variables do not really do the job, which they are expected to do: the separation of method variance from trait variance.

The investigation of discriminant validity in concentrating on the latent variables. There is additionally the possibility to investigate discriminant validity in considering the relationships among the constructs and methods. The consideration of these relationships is of particular importance because the new scale that is to be validated is frequently expected to improve or refine another scale in a particular way and the construction of the new scale has been associated with the modification of the construct. For example, the researcher may state that it is necessary to distinguish between different types of reasoning and to have specific scales for representing these types of reasoning instead of only a general reasoning scale. Therefore, both scales – the old and the new ones – are to be included in the multitrait-multimethod design and special weight is given to the investigation of discriminant validity.

In order to have a guideline for the evaluation of the correlations among the constructs and methods, it was proposed that the standardized correlations between the constructs and also the methods should not surmount the value of .71 at the latent level (Goffin & Jackson, 1992). The value of .71 as limit is expected to prevent cases where two constructs respectively methods have more than 50 % of true variance in common. Correlations between constructs on one hand and methods on the other are not considered since they are set equal to zero.

On the problems of confirmatory factor analysis in multitrait-multimethod data. Unfortunately, the estimation of the parameters of the general model, as it was proposed by Widaman (1985), is very challenging since a large number of parameters have to be estimated. Investigating the general model often leads to estimation problems (see Byrne, in press). A major reason for these problems is the size of the correlations among the scales. Especially the correlations among scales representing the same construct should be very high because otherwise convergent validity is called into question. The consequence is a high probability that there is multicollinearity. So the rank of the multitrait-multimethod matrix may be lower than it is assumed to be, and the degrees of freedom may be an insufficient indicator of the estimability of the parameters of the model. Multicollinearity is also an indication of a restriction of the number of parameters that can be estimated without estimation problems. Since multicollinearity seems to be quite a common problem in this research, it is necessary to be concerned with it explicitly.

Downsizing the complexity of the model is a means that can be very helpful in overcoming the problem. Downsizing can be achieved by fixing correlations that are otherwise estimated to a specific value. It should be applied stepwise and concentrate on correlations that were not estimated appropriately in the full model.

Furthermore, estimation problems may become obvious as negative error variances. Such negative variances indicate that there is an overestimation of the factor loadings leading to the over-extraction of the variances of manifest variables. The construct and method latent variables account for too much variance. Negative error variances also tell the researcher that there is something wrong with the model since normally error variances are considerably larger than zero.

The *ridge option* (Yuan, Wu, & Bentler, 2011) provides a way of dealing with negative error variances and estimation problems in general. The ridge option denotes the procedure that requires the summation of the given correlation matrix with the weighted identity matrix. In a way this procedure means that the variances of all individual manifest variables are increased in a systematic way while the true parts that are common with other manifest variables are kept constant. In this case the ridge option is applied for eliminating negative variances. In contrast to setting negative error variances simply to zero, it can be expected to retain the general structure of the matrix.

Furthermore, there are other models that have been proposed to avoid estimation problems, as for example the *uncorrelated trait-correlated methods (UTCM) model*, the *correlated uniquenesses model* (CUCFA; Kenny, 1979; Marsh, 1988, 1989), and the more recent *correlated trait-correlated methods-minus one model* (CT-C[M-1]; Eid, 2000). However, the CUCFA is only useful as long as the number of assessment methods is small. Otherwise there is a disproportional increase in the number of parameters to be estimated.

An example

For a demonstration of the investigation of discriminant validity two multitrait-multimethod matrices were generated according to the pattern provided by Campbell and

Fiske (1959). One matrix enabled the estimation of the model parameters without estimations problems whereas estimation problems had to be faced in investigating the other one. Since the pattern by Campbell and Fiske was an artificial matrix, it was only used as outset for the construction of simulated random data according to the procedure proposed by Jöreskog and Sörbom (2001). At first, 500 x 9 matrices of continuous and normally distributed random data ($N(0,1)$) were generated. Subsequently, the structure according to the pattern provided by Campbell and Fiske was induced. Finally, 9 x 9 covariance matrices were computed for demonstrating the multitrait-multimethod approach associated with confirmatory factor analysis.

Widaman's (1985) general model provided the outset. This model allowed all the method and all the construct latent variables to correlate with each other with the exception of combinations of constructs and methods. If necessary, the next steps served the reduction of the complexity of this model for achieving a good model-data fit. Subsequently, the complete and modified general models were compared with the models assuming a lack of discrimination between the constructs respectively methods. Finally the correlations of the constructs and methods had to be considered.

The fit statistics are provided in Table 2.

Table 2:
Fit Results Obtained for the Models Considered in Investigating the Discriminant Validity of the Multitrait-multimethod Matrix

Characteristics of the model	Estimation problem	χ^2	df	Normed χ^2	RMSEA	SRMR	CFI	TLI	AIC
Matrix without estimation problems									
CTCM (General)	Yes	11.44	12	0.95	.000	.011	1.00	1.00	77.4
PCTCM	Yes	752.35	15	50.15	.314	.097	0.75	0.39	812.3
CTPCM	No	737.07	15	49.14	.311	.093	0.76	0.42	797.1
Matrix with estimation problems									
CTCM (General)	Yes	14.38	12	1.07	.020	.010	1.00	1.00	80.3
CTCM*	No	22.46	13	1.73	.038	.019	1.00	0.99	86.4
CTCM**	No	18.74	13	1.44	.030	.018	1.00	1.00	82.7
PCTCM	Yes	740.28	15	49.35	.311	.100	0.75	0.40	800.2
CTPCM	No	914.56	16	57.16	.335	.124	0.73	0.40	972.5

Note. CTCM is the acronym of the general model, PCTCM of the general model with the correlations between the trait latent variables fixed to one, and CTPCM of the general model with the correlations between the method latent variables fixed to zero. * indicates that the correlation between the first and third trait latent variables is set equal to one. ** indicates the additional consideration of the ridge option (weight: 0.015).

The first part gives the results for the matrix without estimation problems and the second part for the other matrix. In each part the results for the general model (CTCM) are given in the first row. In the first matrix there were no estimation problems. The fit statistics indicated an excellent model-data fit for the general model and a bad model-data fit for the other models. After scaling the variances of the latent variables of the general model according to Schweizer (2011) in assuming that the average factor loading was .5 the variances of all latent variables turned out to be significant (construct 1: ϕ=1.53, t=6.33, p<.05; construct 2: ϕ=1.37, t=7.25, p<.05; construct 3: ϕ=1.34, t=6.44, p<.05; method 1: ϕ=1.63, t=6.33, p<.05; method 2: ϕ=2.58, t=11.86, p<.05; method 3: ϕ=2.60, t=11.99, p<.05).

Although in the second matrix the model-data fit was also excellent, the program was not able to produce appropriate estimates for the correlations among the construct latent variables. The completely standardized correlations of the constructs varied between -11.49 and -363.30. Furthermore, two error variances were negative (-0.004 and -0.006). As already indicated, this kind of outcome is not that rare in investigating the model-data fit of the general model. There are frequently estimation problems although the degrees of freedom are considerably larger than zero. One source of these problems is the usually strong dependency among the scales in the sense of multicollinarity. In the first step the correlation of which the estimate showed the largest deviation from the range of acceptable values was fixed to zero (CTCM*). As a consequence of this provision, the estimates of the other correlations showed agreeable values. In the second step the ridge option was considered (CTCM**). The weight of 0.015 produced acceptable error estimates. The model-data fit of the modified model was good (see lower part of Table 2). After scaling the variances of the latent variables of the modified model all variances were significant (construct 1: ϕ=1.52, t=10.47, p<.05; construct 2: ϕ=1.80, t=10.47, p<.05; construct 3: ϕ=1.54, t=10.19, p<.05; method 1: ϕ=1.38, t=7.57, p<.05; method 2: ϕ=2.18, t=11.91, p<.05; method 3: ϕ=2.56, t=12.47, p<.05). So in this case the modified general model provided the outset for the investigation of discriminant validity.

In order to demonstrate that there was discrimination between the constructs, the correlations between the three construct latent variables were set equal to one (PCTCM). The model-data fit for this model was very bad in both data matrices. So it was not even necessary to compute the chi-square difference in order to demonstrate that this model differed from the modified general model. Next, the discrimination between the methods was investigated. The correlations between the method latent variables were set equal to one (CTPCM). The model assuming no discrimination between the methods was also characterized by a very bad model-data fit in both data matrices. In both cases it was not really necessary to compute the chi-square difference in order to show that the models differed from each other since the differences were much larger than the critical chi-square differences (7.81 for 3 df and p<.05; 5.99 for 2 df and p<.05). Consequently, there was discrimination between the constructs and the methods.

Finally discriminant validity was investigated in considering the correlations among the construct and method latent variables. In the case of the first data matrix the completely standardized correlations varied between -0.020 and 0.176. All these correlations were insignificant and below the cut-off which was .71. As already indicated, in the case of the second data matrix there was only one substantial correlation between the construct latent variables. The value of this correlation was .24 that was considerably smaller than the cut-off which was .71.

The method latent variables of the first data matrix showed one large correlation of .59 that reached the level of significance but did not surmount the cut-off. The other correlations were .17 and .23. In the second data matrix the correlations between the method latent variables varied between .24 and .54. So there was also no correlation between the method latent variables surmounting the value of .71.

In sum, the fit results concerning the general models, the outcomes of the comparisons of models and of the investigation of the correlations among the construct and method latent variables suggested that there was discriminant validity in both data matrices.

Discussion

Discriminant validity is an important component of construct validity. It is especially well suited for making an unspecific influence of the assessment method on measurement obvious. As is demonstrated by Campbell and Fiske (1959), method variance can boost correlations among scales or scales and criterion variables and thus lead to erroneous conclusions. The investigation of discriminant validity in the framework of a multitrait-multimethod design reduces this danger since it makes the influence of the assessment method obvious and can be employed for separating the construct and method contributions to measurement from each other. In the data matrices selected as examples the scaled variances of the latent variables made obvious that there was a considerable amount of method variance. In the first example the variances of all method latent variables surmounted the other variances, and in the second example two out of three variances of method latent variables were larger than the other variances. Apparently in the two data matrices method variance was a danger to the evaluation of construct validity.

Furthermore, discriminant validity is most suitable for demonstrating the advantage of the new scale that is in the focus of the investigation and normally should mean advancement in the distinction of constructs. Since a major justification for the construction of a new scale is a new construct or the revision of a known construct, the establishment of discriminant validity is an important step in investigating the quality of this new scale. Given a set of already established scales it can be shown that the new scale does something that cannot be accomplished by the available scales. Accordingly, in the two data matrices it was demonstrated that constraining the relationships among the construct latent variables was associated with an impairment of the model-data fit on one hand and that the correlations among these construct latent variables did not surmount the critical value of .71 on the other hand.

However, these advantages of the multitrait-multimethod approach did not really come into operation as long as it was implemented as correlational analysis. It was the implementation in the framework of confirmatory factor analysis that was necessary for eliminating the vagueness, which still characterized the outcomes of investigating multitrait-multimethod matrices. However, clarity concerning discriminant validity can only be established with respect to a few constructs, and it is an open question whether the consideration of these constructs really means a scientific advancement. Discriminant validity established in the framework of a multitrait-multimethod design means that there are a few other constructs, which the scale does not represent. But there is no generalization of discrimination.

Unfortunately, discriminant validity is not achievable without costs. It is because of the establishment of discriminant validity that it is necessary to consider other constructs and to enlarge the design by a multiple of the set of scales selected for the investigation of convergent validity. As a consequence, many comparisons of correlations have to be performed in an investigation of the discriminant validity of a scale by means of the original multitrait-multimethod approach. Confirmatory factor analysis of the multitrait-multimethod matrix improves the situation since it integrates the expectations associated with a multitrait-multimethod matrix into one model and thus simplifies its evaluation. Although the focus of the evaluation is on the model-data fit, in order to be sure concerning the discriminant validity of a scale, it is additionally necessary to take the correlations among the latent variables into consideration.

A disadvantage of the implementation of the multitrait-multimethod approach as confirmatory factor analysis is that it does not distinguish sufficiently well between the scale that is under consideration and the other scales. So in order to establish convergent validity of scale A two other scales B and C may be considered. For investigating discriminant validity of A some more scales are necessary. As a consequence of the necessity of considering a number of other scales besides the scale that is in the focus of the investigation, it may happen that a disadvantageous outcome is due to one or several of the additional scales and not to the scale in the focus. So there is the danger that a bad quality of additional scales may impair the establishment of the discriminant validity of the scale that is in the focus of the investigation.

References

Byrne, B.M. (in press). Testing for evidence of construct validity: the multitrait-multimethod approach. In K. Schweizer & C. DiStefano (Eds.), *Principles and methods of test construction: standards and recent advancements*. Göttingen: Hogrefe.

Byrne, B.M. & Goffin, R.D. (1993). Modeling MTMM data from additive and multiplicative covariance structures: An audit of construct validity concordance. *Multivariate Behavioral Research*, *28*, 67-96.

Campbell, D.T., & Fiske, D.W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*, 81-105.

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

Cizek, G. J., Rosenberg, S. L., & Koons, H. H. (2008). Sources of validity evidence for educational and psychological tests. *Educational and Psychological Measurement, 68*, 397-412.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*, 281-302.

Cudeck, R. (1988). Multiplicative models and MTMM matrices. *Journal of Educational Statistics, 13*, 131-147.

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

Eid, M. (2000). A multitrait-multimethod model with minimal assumptions. *Psychometrika*, *65*, 241-261.

Goffin, R.D., & Jackson, D.N. (1992). Analysis of multitrait-multirater performance appraisal data: Composite direct product method versus confirmatory factor analysis. *Multivariate Behavioral Research*, *27*, 363-385.

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

Jöreskog, K.G., & Sörbom, D. (2001). *Interactive LISREL: User's guide*. Lincolnwood, IL: Scientific Software International Inc.

Kenny, D.A. (1979). *Correlation and causality*. New York: Wiley.

Marsh, H.W. (1988). Multitrait-multimethod analysis. In J.P. Keeves (Ed.), *Educational research methodology, measurement, and evaluation: An international handbook* (pp. 570-578). Oxford: Pergamon.

Marsh, H.W. (1989). Confirmatory factor analyses of multitrait-multimethod data: Many problems and a few solutions. *Applied Psychological Measurement*, *15*, 47-70.

Marsh, H. W. (1989). Confirmatory factor analysis of multitrait-multimethod data: Many problems and a few solutions. *Applied Psychological Measurement, 13*, 335-361.

Norman, W.T., & Goldberg, L.R. (1966). Raters, ratees, and randomness in personality structure. *Journal of Personality and Social Psychology, 4*, 681–691.

Reise, S. P., Morizot, J., & Hays, R. D. (2007). The role of the bifactor model in resolving dimensionality issues in health outcomes measures. *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care & Rehabilitation, 16*, 19-31.

Schweizer, K. (2011). Scaling variances of latent variables by standardizing loadings: applications to working memory and the position effect. *Multivariate Behavioral Research, 46*, 938-955.

Schweizer, K. (2012). On issues of validity and especially on the misery of convergent validity. *European Journal of Psychological Assessment, 28*, 249-254.

Schweizer, K., Altmeyer, M., Reiß, S., & Schreiner, M. (2010). The c-bifactor model as a tool for the construction of semi-homogeneous upper-level measures. *Psychological Test and Assessment Modeling, 52*, 298-312.

Schweizer, K. & Ren, X. (2013). The position effect in tests with a time limit: the consideration of interruption and working speed. *Psychological Test and Assessment Modeling, 55*, 62-78.

Technical recommendations for psychological tests and diagnostic techniques (1954). *Psychological Bulletin Supplement, 51*, Part 2, 1-38.

Tupes, E.C., & Christal, R.E. (1961). Recurrent personality factors based on trait ratings. *Technical Report ASD-TR-61-97*, Lackland Air Force Base, TX: Personnel Laboratory, Air Force Systems Command.

Widaman, K.F. (1985). Hierarchically nested covariance structure models for multitrait-multimethod data. *Applied Psychological Measurement*, *9*, 1-26.

Yuan, K.-H., Wu, R., & Bentler, P. M. (2011). Ridge structural equation modeling with correlation matrices for ordinal and continuous data. *British Journal of Mathematical and Statistical Psychology, 64*, 107-133.

Ziegler, M., Booth, T., & Bensch, D. (2013). Getting entangled in the nomological net. *European Journal of Psychological Assessment, 29*, 157-161.