In press, *British Journal of Mathematical and Statistical Psychology*

Evaluation of Convergent and Discriminant Validity with Multitrait-Multimethod Correlations

Tenko Raykov
Michigan State University

Author Note:

I am indebted to M. D. Reckase and G. A. Marcoulides for valuable discussions on validity estimation, as well as to L. K. Muthén and B. O. Muthén for instructive comments on parameter constraint evaluation. I am grateful to J. Miles and two anonymous Referees for insightful and suggestive criticism on an earlier version of the paper, which contributed substantially to its improvement. Correspondence on this manuscript may be addressed to Tenko Raykov, Michigan State University, Measurement and Quantitative Methods, 443A Erickson Hall, East Lansing, MI 48824, USA; e-mail: raykov@msu.edu .

Running Head: CONVERGENT AND DISCRIMINANT VALIDITY EVALUATION

Abstract

A procedure for evaluation of convergent and discriminant validity coefficients is outlined. The method yields interval estimates of these coefficients in a construct validation study conducted via the multitrait-multimethod approach, as well as permits examination of their population relationships. The procedure is readily employed in behavioral research using the increasingly popular latent variable modeling methodology. The described method is illustrated with a numerical example.

Keywords: construct validity, convergent validity, discriminant validity, interval estimation.

Evaluation of Convergent and Discriminant Validity with Multitrait-Multimethod Correlations

Validity of measurement is of paramount importance in psychology and the behavioral sciences. A crucial indicator of psychometric quality, validity is the bottom line of measurement in these disciplines. A widely accepted informal definition of validity characterizes it as the degree to which an instrument indeed measures what it purports to evaluate (e.g., Crocker & Algina, 1986). For these reasons, validity is an index of critical relevance for a measurement procedure, and has received an impressive amount of attention over the past century (e.g., Borsboom, Mellenbergh, & Van Heerden, 2004).

A major impetus to the study of validity was provided a half century ago by Campbell & Fiske (1959), who introduced the multitrait-multimethod (MTMM) matrix as a means for construct validation. The MTMM method can be used when multiple traits are examined simultaneously and each of them is assessed by a given set of measures or measurement methods (e.g., Eid, 2000; Marsh & Hocevar, 1983). As shown initially by Campbell and Fiske, and further elaborated by subsequent authors, two types of validity coefficients are of special interest when the MTMM matrix is utilized in the validation process—convergent validity and discriminant validity coefficients.

Multitrait-multimethod correlation matrices comprise the (linear) relationship indices among several traits evaluated by different measurement methods. These matrices have been often used over the past few decades by psychologists and behavioral researchers in various substantive areas. Many of those studies, however, have focused on empirical correlations and their relative magnitude in samples from studied populations. Thereby, typically only point estimates of the convergent and discriminant validity coefficients have been of interest. In addition little attention if any has been paid to the population relationships among these coefficients, which are of real concern, and in particular only sample values of these correlations have been usually compared with one another. This has frequently been done in an effort to find out whether discriminant validity coefficients are lower than convergent validity coefficients, a condition posited by Campbell & Fiske (1959) as evidence supporting construct validity. No account has been made then of sampling error affecting the correlation estimates, yet as widely appreciated what is of actual interest for the purpose of construct validation are the relationships between convergent and discriminant validity coefficients at large. For these reasons, incorrect conclusions with respect to validity of behavioral instruments may have been reached in past research. The likelihood of this happening may have been additionally enhanced by the fact that no widely applicable procedure has been made available, which would accomplish examination of these relationships at the population level that is of real relevance for convergent and discriminant validity.

To contribute to closing this gap, the purpose of this article is to outline a readily employed method for (i) interval estimation of convergent and discriminant validity coefficients, as well as (ii) examining their population relationships. The following procedure is straight-forwardly conducted within the framework of latent variable modeling (LVM), and is readily implemented in psychological research with the increasingly popular LVM software M*plus* (Muthén & Muthén, 2008). The goal of the paper is to complement earlier approaches to examining convergent and discriminant validity (e.g., Marsh, 1989; Bagozzi & Yi, 1993). Those approaches are based on confirmatory factor analysis and latent variables that represent studied traits as well as somewhat more difficult to interpret 'method factors'. In difference to them, the following method is developed in terms of observed variables. More specifically, it is focused on the relationships between collected measures in an empirical study, and in this way circumvents possible substantive interpretation difficulties associated with 'method factors'. To accomplish these aims, in the remainder of this article we use the analytic technique of latent variable modeling as a means for achieving interval estimation of manifest measure correlations, rather than for fitting models based on latent variables evaluated with multiple indicators.

Point and Interval Estimation of Convergent and Discriminant Validity

Convergent and discriminant validity coefficients

In the context of an application of the MTMM method for construct validation, as discussed by Campbell & Fiske (1959) the convergent validity coefficients are the correlations between measures of the same trait that are obtained with different measurement methods. For this reason, those correlations are at times referred to as monotrait-heteromethod (MTHM) coefficients. Since they reflect the (linear) relationships between indicators of the same trait, a finding of them being consistently high lends support for construct validity with regard to that trait. The evaluation of these coefficients thus represents part of the validation process with the MTMM method.

Conversely, discriminant validity coefficients are correlations of two types. The first comprises the correlations between measures of different traits that are furnished by the same method of measurement, and are called heterotrait-monomethod coefficients (HTMM coefficients). The second type consists of correlations between measures of different traits that are obtained using different measurement methods, and are called heterotrait-heteromethod coefficients (HTHM coefficients). A finding that the HTMM and HTHM coefficients are consistently lower than the MTHM coefficients lends also support for validity with regard to each of the traits involved, since one would usually expect indicators of a given construct to be more closely related than measures of different constructs. Hence, examination of the HTMM and HTHM correlations represents another part of the validation process with the MTMM approach.

Estimation of validity coefficients

Underlying model

In order to point and interval estimate these three types of validity coefficients, a special confirmatory factor analysis (CFA) model can be used in a first step (Kühnel, 1988). This model differs from popular CFA approaches to examining convergent and discriminant validity, which have been available for more than two decades (e.g., Bagozzi & Yi, 1993; Marsh & Hocevar, 1983; Marsh, 1989; Eid, 2000), in one main feature. This is the fact that in the cited procedures use is made of ‘proper’ latent variables that are evaluated by multiple indicators. The latter are the collected multiple measures of several traits using several methods of evaluation, whereas the latent variables represent the traits themselves and somewhat more difficult to interpret latent constructs. These constructs have been included in part to explain residual variation among measures of the same trait with different modes of assessment, and have been frequently referred to as ‘method factors’ (e.g., Marsh & Hocevar, 1983).

Unlike those earlier approaches, the present article does not use proper latent variables and remains at the observed variable level in its analytic endeavor. For our purposes, let us denote first by $\underline{X}$ the p x 1 vector of available measures on all c traits involved in an MTMM study (p, $c \geq 2$); that is, $p = mc$ is the product of the number c of traits with that of measurement methods (underlining is used to denote vector in the remaining discussion and priming for transposition). Let $\underline{X}_1 = (X_1, \ldots, X_c)'$, $\underline{X}_2 = (X_{c+1}, \ldots, X_{2c})'$, …, $\underline{X}_m = (X_{(m-1)c+1}, \ldots, X_p)'$, and hence $\underline{X} = (\underline{X}_1' | \underline{X}_2' | \ldots | \underline{X}_m')'$, where '|' denotes concatenation. Then the model of interest in the rest of this article is

(1) $\underline{X} = \mathrm{A}\,\underline{\mathrm{f}} + \underline{\mathrm{u}}$,

where $\mathrm{A} = \mathrm{I}_{pxp}$ is a diagonal matrix of size p with main diagonal elements being model parameters, $\underline{\mathrm{f}}$ is a set of p dummy 'latent' variables with variance 1 whose covariances are also model parameters, and $\underline{\mathrm{u}}$ is a zero-mean vector with zero covariance matrix, i.e., consisting of zeroes only. We stress that each of the 'latent' variables in $\underline{\mathrm{f}}$ is formally identical to a corresponding observed variable in $\underline{X}$, since each of the former is measured by a single observed variable and associated with no measurement error. The 'latent' dummy variables in $\underline{\mathrm{f}}$ will be used shortly for accomplishing the specific goals of this article.

Convergent and discriminant validity coefficient estimation

With this particular parameterization of the model in Equation (1), obviously each observed measure, X_i, is identical to its corresponding latent variable, f_i, up to a multiplication constant that equals that measure's standard deviation, i.e.,

(2) $X_i = \sigma_i \, . \, \mathrm{f}_i$,

where σ_i denotes the standard deviation of X_i ($i = 1, \ldots, p$). Hence, the covariance matrix of the random vector $\underline{X}$ is the same as its correlation matrix, i.e.,

(3) $Cov(\underline{\mathrm{f}}) = Corr(\underline{\mathrm{f}})$,

where *Cov*(.) and *Corr*(.) denote covariance and correlation matrix of random vector within parentheses, respectively.

Equations (1) through (3) imply that point and interval estimation of the correlations among elements of the vector $\underline{f}$ is accomplished by point and interval estimation of the corresponding covariances among them. Indeed, due to Equation (2) and the well-known invariance property of the correlation coefficient to linear transformation of involved variables, it is easily realized that the matrix $Corr(\underline{f})$ consists of the above convergent and discriminant validity correlations positioned at appropriate places. To give a few examples of relevance later in the illustration section of this article, $Corr(f_1, f_{c+1})$ is the MTHM (convergent validity) correlation between the measures of the first listed trait that are obtained with the first and second listed measurement methods. Similarly, $Corr(f_2, f_{c+1})$ is the HTHM (discriminant validity) correlation between the measures of the first two traits that are obtained with the first two measurement methods. Also, $Corr(f_1, f_2)$ is the HTMM (discriminant validity) correlation of the measures of the first and second traits that have been obtained with the first measurement method.

<u>Differences from prior approaches to studying convergent and discriminant validity</u>

The modeling approach underlying this article complements earlier procedures for examining convergent and discriminant validity (e.g., Bagozzi & Yi, 1993; Marsh & Hocevar, 1983; Marsh, 1989; Eid, 2000). This is due to an essential feature that is seen from the defining Equation (1). Specifically, this approach does not contain error terms, unlike those earlier methods where each observed variable is associated with such a residual. The reason is that here we are interested in the study of the relationships of convergent and discriminant validity coefficients that are defined in terms of observed measure correlations (e.g., Campbell & Fiske, 1959; cf., Crocker & Algina, 1986; McDonald, 1999). In the CFA-based approaches, these correlations were primarily used as a starting point for the estimation of trait correlations, which was in part achieved by the inclusion of measurement error parameters (error variances). There are no such parameters in the present method, owing to its goal being interval estimation of convergent and discriminant validity coefficients in observed variable correlations. For this reason, the present approach does not contain any parameters for the trait interrelationships, and does not provide any information about them. Hence, it cannot be used when the trait interrelationships are of interest; in that case, appropriate from those CFA-based approaches can be utilized (e.g., Bagozzi & Yi, 1993; Marsh & Hocevar, 1983; Marsh, 1989; Eid, 2000).

Latent variable modeling as a technique for interval estimation of convergent and discriminant validity coefficients

Equation (1) can be looked at as formally defining a latent variable model since it contains latent variables and measures of them (e.g., Muthén, 2002). Since there are no restrictions imposed within it, which would have implications for its covariance structure, this model is saturated. For this reason, it fits a given *pxp* covariance matrix perfectly (regardless whether it stems from a studied population or a sample). In an empirical setting, this model can be fitted to data with widely circulated software, such as LISREL, EQS, R, or M*plus* (see below). Resulting from this model fitting process are parameter estimates and standard errors for each of the convergent and discriminant validity coefficients of relevance in the context of the MTMM approach. When maximum likelihood estimation is used, these estimates are the same as the corresponding elements of the sample correlation matrix.

While estimation of these correlations is desirable in its own right, this is not the main benefit from an application of the present method. Specifically, we stress that when fitting the underlying CFA model defined in Equation (1) we obtain also standard errors for the observed measure correlations, which standard errors will play instrumental role next. This feature of the method is not shared with the above mentioned prior procedures to discriminant and convergent validity examination, and it is in this and related properties that the method in this article complements those procedures.

While the so-obtained MTMM correlation estimates represent optimal numerical ‘guesses’ for the population correlations between the trait measures under consideration, the estimates do not contain any information as to how far they may be from their counterpart ‘true’ correlations in the population that are of actual interest in a validation study. As a major goal of the present method, such information can be obtained in the form of a confidence interval (CI) via the following procedure.

Confidence interval construction for validity coefficients

In order to accomplish interval estimation of convergent or discriminant validity coefficients, we make first an important observation. Each of these coefficients, being a correlation, is bounded by -1 and 1; therefore, a CI for it cannot contain values that are larger than 1 or smaller than -1. In fact, for typical convergent or discriminant validity coefficients in empirical behavioral research, one can argue that the values of 1 and -1 are not plausible either, like values very close to them. Hence, it would be also justified to expect that a meaningful CI for such a validity coefficient would typically not include 1 or -1, and not necessarily values in their closest proximity. This expectation is further corroborated by the well-known fact (e.g., Agresti & Finlay, 2008) that the sampling distribution of a correlation coefficient is in general non-symmetric, and specifically skewed to the right/left in case of positive/negative population correlation.

With this in mind, a large sample confidence interval for a convergent or discriminant validity coefficient can be obtained by initially furnishing such a confidence interval for an appropriate monotonically increasing transformation of the coefficient, such as Fisher's *z*-transformation (e.g., Browne, 1982). That is, denoting a convergent or discriminant correlation generically by ρ, its transformed value is

(4) $$z(\rho) = .5ln[(1+\rho)/(1-\rho)] \,,$$

where $ln[.]$ denotes natural logarithm. Then a 95%-CI for the transformed population coefficient, $z(\rho)$, can be obtained as well known by subtracting and adding 1.96 times the associated standard error to its estimate $z(\hat{\rho})$ furnished via Equation (4) using the correlation estimate $\hat{\rho}$ resulting when fitting model (1) to data. This standard error, denoted $SE(z(\hat{\rho}))$, can be shown to equal

(5) $$SE(z(\hat{\rho})) = SE(\hat{\rho})/(1-\hat{\rho}^2)$$

(Browne, 1982), where $SE(\hat{\rho})$ is the standard error of the estimate $\hat{\rho}$ of the convergent/ discriminant validity coefficient of concern. (As mentioned above, $SE(\hat{\rho})$ is also obtained, along with $\hat{\rho}$, when fitting model (1).) That is, the 95%-CI for the transformed convergent/discriminant validity coefficient, $z(\rho)$, is

(6) $(z(\hat{\rho})\text{-}1.96 \text{ x } SE(\hat{\rho})/(1\text{-} \hat{\rho}^2); z(\hat{\rho})\text{-}1.96 \text{ x } SE(\hat{\rho})/(1\text{-} \hat{\rho}^2))$,

where 'x' symbolizes multiplication. Finally, the sought approximate 95%-CI of the population convergent/discriminant validity coefficient ρ of actual concern is obtained by using the inverse of Fisher's z-transformation, viz.

(7) $$h(z) = \frac{\exp(2z)-1}{\exp(2z)+1} ,$$

where exp(.) denotes exponent of the quantity in parentheses. That is, the lower and upper limits of the required CI result by applying the inverse Fisher transformation on the corresponding endpoints of the CI in (6):

(8) $((\exp(2z_{lo}) - 1)/(\exp(2z_{lo}) + 1); (\exp(2z_{up}) - 1)/(\exp(2z_{\rho,up}) + 1))$,

where z_{lo} and z_{up} denote the left and right limits of (6), respectively. [1] (In general, for a confidence level of 1-α, one uses in (6) the corresponding cutoff $z_{\alpha/2}$ of the standard normal distribution in lieu of 1.96; $0 < \alpha < 1$.) Computation of the endpoints of (8) is readily accomplished in empirical research with popular LVM software as described in a later section.

Studying Population Relationships Between Convergent and Discriminant Validity Coefficients

Comparing convergent and discriminant validity coefficients

As indicated earlier, findings in a psychological study that convergent validity coefficients are larger than discriminant validity coefficients lend support to construct validity when using the MTMM approach. The latter would be the case, however, when this type of relationship holds in the population of concern (as opposed to holding in a sample only). A comparison of confidence intervals of the corresponding coefficients, obtained as in the preceding section, permits drawing some informal conclusions about these population relationships. However, due to these CIs resulting from the same set of analyzed data, they are related to one another. Hence, even in cases when the CIs for a convergent and discriminant validity coefficient do not overlap, one cannot strictly use them to ascertain the relationship between their population values of actual relevance.

This population relationship can be examined using a confidence interval for the difference of two validity coefficients, e.g., a convergent and a discriminant coefficient, denoted next by ρ_1 and ρ_2. (The same procedure can be used when of interest is to compare the population values of the two types of discriminant validity coefficients, such as HTMM and HTHM correlations; a finding of the former being larger provides support for construct validation, e.g., Campbell & Fiske, 1959.) At a given confidence level, this difference CI can be obtained using the generally applicable delta method (e.g., Roussas, 1997), which will also be used later for more complex functions of MTMM correlations. To this end, symbolize by $D(\underline{\rho})$ their difference, i.e., $D(\underline{\rho}) = \rho_1 - \rho_2$, where $\underline{\rho} = (\rho_1, \rho_2)'$. (We note in passing that these two correlations are parameters of the model defined by Equation (1).) The first-order Taylor expansion of $D(\underline{\rho})$, as a function of the two correlations involved, around the corresponding population parameter $\underline{\rho}_0$ is (e.g., Stewart, 1991):

$$(9) \qquad D(\underline{\rho}) \approx D(\underline{\rho}_0) + \left[\frac{\partial D(\underline{\rho})}{\partial \underline{\rho}'} \Big|_{\underline{\rho}=\underline{\rho}_0} \right] (\underline{\rho} - \underline{\rho}_0) \ ,$$

where the symbol '≈' denotes 'approximately equal' and bracketed is the vector of partial derivatives of D($\underline{\rho}$) with respect to the two correlations it depends on (which derivatives are 1 and -1 in this case, but need not be so in a more general case as further below). From Equation (9), a squared large-sample standard error of the difference between a convergent and discriminant validity coefficients follows as

$$(10) \qquad Var(D(\hat{\rho})) \approx \left[\frac{\partial D(\hat{\underline{\rho}})}{\partial \underline{\rho}'} \Big|_{\underline{\rho}=\hat{\underline{\rho}}} \right] Cov(\hat{\underline{\rho}}) \left[\frac{\partial D(\hat{\underline{\rho}})}{\partial \underline{\rho}} \Big|_{\underline{\rho}=\hat{\underline{\rho}}} \right],$$

where $Cov(\hat{\underline{\rho}})$ is the relevant 2 x 2 part of the observed inverted information matrix associated with the fitted model, which pertains to the two correlation parameters in question. Using Equation (10), a large-sample 100(1 - α)%-confidence interval (0 < α < 1) for the convergent-discriminant validity coefficient difference is readily furnished as

$$(11) \qquad D(\hat{\underline{\rho}}) \pm z_{\alpha/2}\sqrt{VarD(\hat{\rho})} \; ,$$

where $z_{\alpha/2}$ is the (1-α/2)th quantile of the standard normal distribution. The limits of this confidence interval are also easily obtained using popular LVM software as described in a following section.

Evaluation of overall difference between convergent and discriminant validity coefficients

The outlined interval estimation approach allows one to compare the population values of any pair of convergent and discriminant validity coefficients (or any pair consisting only of convergent coefficients, or only of discriminant validity coefficients). The method is best used on a limited number of pairs of validity coefficients, however, due to well-known problems with multiple inferences carried out on the same data set (e.g., Agresti & Finlay, 2008). However, in extensive MTMM studies there may be a fairly large number of pairs of interest to examine. In addition, and at least as importantly, it may be also desirable to obtain an overall statement as to whether convergent validity coefficients are uniformly, or tend to be, larger than discriminant validity coefficients (and similarly with respect to HTMM and HTHM discriminant validity correlations; Campbell & Fiske, 1959). In such cases, the outlined delta-method based procedure can also be applied, so long as one pre-specifies an appropriate function of the two types of validity coefficients that are to be compared at the population level.

As such a function, it may be argued that the mean validity coefficient could be frequently of interest in psychological research. Accordingly, a question to be addressed is whether the average convergent validity coefficient is larger than the average discriminant coefficient in the studied population. This query can be responded to by applying the last discussed CI-construction procedure, where $D(\rho)$ is now chosen as the difference between these two means, i.e.,

$$\text{(12)} \qquad D(\rho) = \frac{\rho_{11} + \rho_{12} + ... + \rho_{1a}}{a} - \frac{\rho_{11} + \rho_{12} + ... + \rho_{1b}}{b} \;,$$

with ρ_{11} through ρ_{1a} symbolizing the convergent validity coefficients (or any subset of a of them that is of interest) while ρ_{11} through ρ_{1b} denote the discriminant validity coefficients (or a subset of b of them that are of concern in such a comparison).

The desired CI is then provided by Equations (10) and (11) above, where $D(\rho)$ from Equation (12) is formally substituted. (Its derivatives with regard to the correlations are readily obtained using standard differentiation rules, but are actually not needed here to use explicitly, due to the delta method being implemented in the LVM software utilized later; see next section.) If the resulting CI for $D(\rho)$ is entirely located above 0 one can interpret the finding as suggestive, with high confidence, that on average convergent validity coefficients are higher than discriminant validity coefficients. Such a result would provide support for construct validation. If this CI contains the 0 point, then it would be suggested that on average the former validity coefficients do not differ from the latter validity coefficients. In such cases, an important condition may not be fulfilled in the studied population, which according to Campbell & Fiske (1959) would be expected to hold in case of high construct validity. This may cast potentially serious doubts with respect to lacking construct validity regarding some of the traits involved in the study.

This confidence interval procedure can also be used, in the same way, if one wishes to compare the mean HTMM correlations with the mean HTHM correlations in the population, where a finding of the former being larger may be of interest in the context in a validation study (e.g., Crocker & Algina, 1986). This is accomplished by substituting these correlation coefficients correspondingly into the right-hand side of Equation (12) and proceeding as outlined above.

We discuss next the empirical computation of the endpoints of the confidence intervals described in this article, and illustrate in on data in the following section.

Application of Outlined Interval Estimation Procedures in Empirical Research

The confidence interval construction method outlined above is readily employed in a behavioral study using the increasingly popular LVM software M*plus* (Muthén & Muthén, 2008). To this end, in the first step one fits the CFA model in Equation (1), introducing p dummy latent variables, i.e., as many such variables as there are observed measures (viz. $p = cm$). This is achieved by defining each observed variable as loading on only one latent variable, whose variance is constrained at 1; its loading on the measure is at the same time a model parameter. In this way, estimates with standard errors of the convergent and discriminant validity coefficients are obtained. In the second step, one introduces parametric symbols for each correlation involved in a population comparison of interest to conduct, and adds a MODEL CONSTRAINT section. In the latter, initially defined as new parameters are all necessary quantities for obtaining the limits of the confidence intervals of concern (e.g., (8) or (11) for an appropriately chosen correlation function D(ρ); see Appendix for source code details). To obtain a CI for any convergent or discriminant validity coefficient, the CI endpoints are calculated by implementing their formulas in Equation (11) using the corresponding parameter symbols. For interval estimation of the difference in two correlations of interest, or of the mean difference (12) of convergent and discriminant validity coefficients (or any other function of such that could be of interest), one defines formally this difference or function involving the corresponding parameters. All these activities are incorporated in the software source codes provided in the Appendix, which are used there on data from the example considered next (see also notes to codes).

Illustration on Data

To demonstrate the utility and applicability of the described interval estimation procedure, consider the following example. Suppose we were interested in measuring guilt by self-report inventories (cf. Crocker & Algina, 1986, and references therein). We focus on the constructs of hostility guilt and morality conscience, and assume they are measured by three methods: true-false, forced-choice, and incomplete sentences tests (scales). The correlation matrix, standard deviations, and means of these 6 observed measures from $n = 435$ cases, for which the assumption of multinormality can be viewed as plausible, are presented in Table 1.

Insert Table 1 about here

As a first step in the application of the estimation procedure in this article, the CFA model defined in Equation (1) is fitted to these data using the popular LVM program M*plus* (see source code for Step 1 in Appendix, as well as note to it). As mentioned earlier, this model is saturated and hence associated with perfect fit: chi-square = 0, df = 0, root mean square error of approximation = 0 with a 90%-confidence interval (0, 0). A particular benefit of fitting the model lies as indicated in obtaining with it standard errors for the convergent and discriminant validity coefficients, which are of relevance next.

Suppose one were interested in evaluating the population discriminant validity coefficient of the true/false tests for the hostility guilt and morality conscience traits, i.e., the correlation $Corr(f_1,f_2)$. From Table 1, its estimate is .45, while from the results of Step 1 its standard error is obtained as .038. To interval estimate this validity coefficient, we substitute these quantities into Equations (4) through (8), yielding (.371, .521) as the 95%-CI for its population value. These computational activities are straight-forwardly accomplished using the source code for Step 2 presented in the Appendix (see also note to it). This result suggests, with high confidence, that the discriminant validity coefficient in question may lie in the population between the high .30s and high .50s. Since this is an HTMM correlation, such a finding is not unexpected as these correlations are usually not very high or very low.

Alternatively, suppose one were interested in obtaining a CI of the convergent validity coefficient for the true-false and incomplete sentences measures of the construct hostility guilt, i.e., the correlation *Corr*(f_1,f_3). From Table 1, this correlation is estimated at .85, and from the results of Step 1 we find its standard error as .013. Entering these two quantities correspondingly in the source code for Step 2, yields the 95%-CI for this convergent validity coefficient as (.822, .874). This finding suggests, with high degree of confidence, that the population convergent validity coefficient for these measures of the hostility guilt trait is in the mid .80s. Like the previous result, such a finding could also be expected from different indicators (measures) of the same construct, as we deal with here.

Assume next that we wanted to see whether in the studied population a convergent validity coefficient is larger than a discriminant validity coefficient, as one would expect on theoretical grounds. To this end, suppose one were interested in examining if say the population correlation ρ_3 between the true/false and incomplete sentence measures of the hostility guilt construct (i.e., *Corr*(f_1,f_5)) is larger than the correlation ρ_4 between the true/false measures of this construct with that of morality conscience (i.e., *Corr*(f_1,f_2)). To address this question, constructing a CI of the correlation difference $D(\underline{\rho}) = \rho_3 - \rho_4$ is appropriate, which is achieved using Equation (11). The computation of its lower and upper endpoints is readily carried out with the delta method that is implemented in M*plus* and automatically invoked by the software upon request (see source code for Step 3 in Appendix, and note to it.) The resulting 95%-CI for the difference in these convergence and discriminant validity coefficients is (.108, .292). This result suggests, with high confidence, that the convergence validity coefficient considered is markedly higher in the population than the discriminant validity coefficient in question. Such a finding would also be consistent with theoretical expectations, given that the convergence validity coefficient, ρ_3, is a correlation between two indicators of the same construct (an MTHM correlation); at the same time, the discriminant validity coefficient, ρ_4, is a correlation between same method measures of different constructs (an HTMM correlation), and thus could be expected to be considerably lower in the population.

The last illustration was concerned with a comparison of population values of only two validity coefficients while, as mentioned earlier, when using the MTMM approach in validation studies one is usually interested in whether the convergent validity coefficients exhibit an overall tendency of being higher than the discriminant validity coefficients. With this in mind, in the current example it would be of interest to know whether the mean of the 6 convergent validity coefficients (i.e., the MTHM correlations that are positioned along the second diagonal of the matrix in Table 1) is higher in the population than the mean of the 9 discriminant validity coefficients (viz. the HTMM and HTHM correlations, i.e., the remaining correlations). Such a comparison can be made using the CI for the difference in the two mean validity coefficients, as elaborated in the preceding section (see Equations (11) and (12)). The computation of the endpoints of this CI is accomplished with the last source code in the Appendix (see also note to it), and the resulting 95%-CI is (.343, .468). This suggests, with high confidence, that in the population the convergence validity coefficients exceed on average the discriminant validity coefficients by an amount that could be as low as .343 and as high as .468. This finding is consistent with theoretical expectations, given that convergent validity coefficients reflect relationships between different measures of the same traits whereas discriminant validity coefficients reflect considerably weaker relationships between different indicators of different traits.

Conclusion

This article outlined a readily applicable procedure for point and interval estimation of convergent and discriminant validity coefficients that play a major role in validation studies in psychological and behavioral research. The procedure complements previous approaches to the study of convergent and discriminant validity (e.g., Bagozzi & Yi, 1993; Marsh & Hocevar, 1983; Marsh, 1989; Eid, 2000), which were developed in terms of 'proper' latent variables with fallible multiple indicators, the available observed measures. Those CFA-based approaches were concerned with the relationships between underlying traits and possible method factors, as well as their relationships with the manifest measures. In difference to them, the present note was concerned with the relationships between the manifest measure correlations, and specifically examined relationships among discriminant and convergent validity coefficients (Campbell & Fiske, 1959; Crocker & Algina, 1986). In addition, we note that the procedure in this article is equally applicable with missing data, under the assumption of multinormality and missing at random, using the full information maximum likelihood method (e.g., Little & Rubin, 2002). Last but not least, the underlying method in this paper for interval estimation of correlation coefficients and functions thereof, is applicable also with the previous CFA-based approaches for the purpose of interval estimation of trait or other latent correlations and functions of them while allowing measurement error to be taken into account (cf. Browne, 1982).

Several limitations of the outlined procedure need to be noted. Due to being based on an application of the LVM methodology that rests upon an asymptotic statistical theory, the present method is best used with large samples. (We note in passing that this limitation is shared with the earlier, CFA-based approaches mentioned in the preceding paragraph, as the latter also use instrumentally the same methodology.) Further, since the invariance property of maximum likelihood is utilized in some applications of the present procedure (viz. differences between validity coefficients), they are optimally used with multivariate normality. However, robust maximum likelihood estimation can be used for interval estimation of individual validity coefficients with violations of multivariate normality that do not result from piling of cases at extreme ends of underlying scales, marked clustering effects or highly discrete observed measures (Muthén & Muthén, 2008). Moreover, as indicated earlier in the paper, its method cannot be used if interest lies in evaluating underlying trait interrelationships, since there are no parameters in it that reflect them.

In conclusion, this article provides psychologists and behavioral researchers with a readily utilized method for obtaining ranges of plausible population values for convergent and discriminant validity coefficients, as well as differences among them and important functions of them. With this feature, the method complements earlier approaches to the study of convergent and discriminant validity that were developed within the framework of confirmatory factor analysis. In this way the present procedure, along with those earlier approaches, can further aid significantly psychologists' construct validation efforts with regard to measures developed for studying populations of interest.

References

Agresti, A. & Finlay, B. (2008). *Statistical methods for the social sciences*. Upper Saddle River, NJ: Prentice Hall.

Bagozzi, R. P. & Yi, Y. Y. (1993). Multitrait-multimethod matrices in consumer research: Critique and new developments. *Journal of Consumer Psychology, 2,* 127-142.

Borsboom, D., Mellenbergh, G. J., & Van Heerden, J. (2004). The concept of validity. *Psychological Review, 111*, 1061-1071.

Browne, M. W. (1982). Covariance structures. In D. M. Dawkins (Ed.), *Topics in applied multivariate analysis* (pp. 72-141). Cambridge: Cambridge University Press.

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

Crocker, L. S., & Algina, J. (1986). *Introduction to classical and modern test theory.* Forth Worth, TX: Harcourt Brace Jovanovich.

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

Kühnel, S. (1988). Testing MANOVA Designs with LISREL. *Sociological Methods and Research*, *16*, 504-523.

Little, R., & Rubin, D. B. (2002). *Statistical analysis with missing data.* New York: Wiley.

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

Marsh, H. W., & Hocevar, D. (1983). Confirmatory factor analysis of multitrait-multimethod matrices. *Journal of Educational Measurement, 20,* 231-248.

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

Muthén, B. O. (2002). Beyond SEM: General latent variable modeling. *Behaviormetrika, 29*, 87-117.

Muthén, L. K., & Muthén, B. O. (2008). *MPLUS user's guide*. Los Angeles, CA: Muthén & Muthén.

Roussas, G. G. (1997). *A course in mathematical statistics*. New York: Academic Press.

Stewart, J. C. (1991). *Calculus*. Monterey, CA: Brooks/Cole.

Footnote

[1] Due to Fisher's *z*-transformation being a nonlinear function, the confidence interval in (8) is in general not symmetric. While one can obtain a symmetric confidence interval using the normal approximation for the maximum likelihood correlation estimates, since the population correlation is a bounded parameter such an interval will not be optimal in general (e.g., Browne, 1982).

Table 1. Correlations among 6 measures of 2 constructs, resulting from 3 measurement methods, with standard deviations (underneath) and means (bottom line; *n* = 435).

Measure	M1T1	M1T2	M2T1	M2T2	M3T1	M3T2
M1T1	1					
M1T2	.45	1				
M2T1	.85	.40	1			
M2T2	.35	.80	.30	1		
M3T1	.65	.15	.70	.25	1	
M3T2	.20	.55	.20	.65	.35	1
StdDev:	202.35	331.26	298.35	255.67	322.78	301.34
Means:	42.25	48.33	52.35	45.23	49.31	51.86

Note. M*i*T*j* = Measure with the *i*th method of the *j*th trait (*i* = 1, 2, *j* = 1, 2, 3); StdDev = standard deviation; *n* = sample size.

Appendix 1

M*plus* Source Codes for Interval Estimation of Convergent and Discriminant Validity Coefficients and their Relationships

```
TITLE:      MULTITRAIT-MULTIMETHOD MATRIX BASED CONSTRUCT VALIDATION USING
            LATENT VARIABLE MODELING.  STEP 1 (SE for a validity coefficient)
            (SEE ANNOTATING COMMENTS AFTER EXCLAMATION MARK IN PERTINENT ROW)
DATA:       FILE = TABLE1.DAT; ! See Table 1.
            TYPE = CORRELATION MEANS STDEVIATIONS;
            NOBSERVATIONS = 435;
VARIABLE:   NAMES = M1T1 M1T2 M2T1 M2T2 M3T1 M3T2; ! See note to Table 1.
MODEL:      F1 BY M1T1*1; M1T1@0;
            F2 BY M1T2*1; M1T2@0;
            F3 BY M2T1*1; M2T1@0;
            F4 BY M2T2*1; M2T2@0;
            F5 BY M3T1*1; M3T1@0;
            F6 BY M3T2*1; M3T2@0;
            F1-F6@1;
```

Note. The TITLE command provides a succinct description of the goal of the analysis, and the DATA command directs the software to the data to be analyzed. The VARIABLE command assigns names to the variables. The MODEL command section defines each of the 6 dummy latent variables, denoted F1 through F6, as identical to the corresponding observed measure up to a constant (with associated error term set at 0; e.g., Kühnel, 1988). The variances of all latent variables are fixed at 1 (last line), in order (1) to achieve model identification, and (2) to ensure the instrumental identity of latent correlations to corresponding covariances (see main text).

```
TITLE:      MULTITRAIT-MULTIMETHOD MATRIX BASED CONSTRUCT VALIDATION USING
            LATENT VARIABLE MODELING.  STEP 2 (CI for a validity coefficient)
            (SEE ANNOTATING COMMENTS AFTER EXCLAMATION MARK IN PERTINENT ROW)
DATA:       FILE = TABLE1.DAT;
            TYPE = CORRELATION MEANS STDEVIATIONS;
            NOBSERVATIONS = 435;
VARIABLE:   NAMES = M1T1 M1T2 M2T1 M2T2 M3T1 M3T2;
MODEL:      F1 BY M1T1*1; M1T1@0;
```

```
            F2 BY M1T2*1; M1T2@0;
            F3 BY M2T1*1; M2T1@0;
            F4 BY M2T2*1; M2T2@0;
            F5 BY M3T1*1; M3T1@0;
            F6 BY M3T2*1; M3T2@0;
            F1-F6@1;
MODEL CONSTRAINT:
            NEW(RHO, SE, Z, SEZ, CI_Z_LO, CI_Z_UP, CI_C_LO, CI_C_UP);
            RHO = .45; ! ENTER THE CORRELATION ESTIMATE FROM STEP 1.
            SE = .038; ! ENTER ITS ASSSOCIATED S.E., FROM STEP 1.
            Z = .5*LOG((1+RHO)/(1-RHO)); ! FISHER'S Z-TRANSFORM OF RHO.
            SEZ = SE/(1-RHO**2); ! THIS IS ITS PERTINENT S.E.
            CI_Z_LO = Z-1.96*SEZ; ! CI FOR FISHER'S Z-TRANSFORM OF RHO.
            CI_Z_UP = Z+1.96*SEZ; ! (SEE EQUATION (7)).
            CI_C_LO = (EXP(2*CI_Z_LO)-1)/(EXP(2*CI_Z_LO)+1); ! SEE EQ. (8)
            CI_C_UP = (EXP(2*CI_Z_UP)-1)/(EXP(2*CI_Z_UP)+1);
```

Note. This command file is identical the one for Step 1 in its 14 lines, used here as first 14 lines as well, and extends it with the MODEL CONSTRAINT section. The latter achieves the computational activities needed for the construction of the CI of the validity coefficient of interest. After introducing first (with the NEW subcommand) all quantities needed, the following equations mirror Equations (4) through (8). (One needs to enter the coefficient estimate, .45 here, and standard error, .038 here, which are obtained with the preceding command file). This set of 8 equation srenders the lower and upper endpoint of the sought 95%-confidence interval in the quantities CI_C_LO and CI_C_UP. (Use a correspondingly modified cutoff, in lieu of 1.96, to obtain a CI at another confidence level.)

```
TITLE:      MULTITRAIT-MULTIMETHOD MATRIX BASED CONSTRUCT VALIDATION USING
            LATENT VARIABLE MODELING. STEP 3 (CI for diff. in val. coeff.)
            (SEE ANNOTATING COMMENTS AFTER EXCLAMATION MARK IN PERTINENT ROW)
DATA:       FILE = TABLE1.DAT;
            TYPE = CORRELATION MEANS STDEVIATIONS;
            NOBSERVATIONS = 435;
```

```
VARIABLE:   NAMES = M1T1 M1T2 M2T1 M2T2 M3T1 M3T2;
MODEL:      F1 BY M1T1*1; M1T1@0;
            F2 BY M1T2*1; M1T2@0;
            F3 BY M2T1*1; M2T1@0;
            F4 BY M2T2*1; M2T2@0;
            F5 BY M3T1*1; M3T1@0;
            F6 BY M3T2*1; M3T2@0;
            F1-F6@1;
            F1 WITH F5(RHO3); ! CONVERGENT VALIDITY COEFFICIENT OF INTEREST
            F1 WITH F2(RHO4); ! DISCRIMINANT VALIDITY COEFF. OF INTEREST
MODEL CONSTRAINT:
            NEW(DELTA);
            DELTA = RHO3 – RHO4; ! DIFFERENCE OF VALIDITY COEFFICIENTS
OUTPUT:     CINTERVAL; ! REQUESTS CI OF CONCERN.
```

Note. This command file differs from that for Step 1 only in its MODEL CONSTRAINT section. In the latter, initially the quantity DELTA is introduced as a new parameter and defined as the difference of the 2 validity coefficients of interest, assigned the parametric symbols RHO3 and RHO4. In difference to the command file in Step 1 (and Step 2), the present uses in its MODEL CONSTRAINT section only the 2 lines needed for defining the correlation difference of interest, and the OUTPUT statement requesting its confidence interval evaluation.

```
TITLE:      MULTITRAIT-MULTIMETHOD MATRIX BASED VALIDATION
            STEP 4 (CI for mean difference in convergent and discriminant
            validity coefficients)
DATA:       FILE = TABLE8_6.DAT; ! see Table 8.6
            TYPE = CORRELATION MEANS STDEVIATIONS;
            NOBSERVATIONS = 435;
VARIABLE:   NAMES = M1T1 M1T2 M2T1 M2T2 M3T1 M3T2;
MODEL:      F1 BY M1T1*1; M1T1@0;
            F2 BY M1T2*1; M1T2@0;
            F3 BY M2T1*1; M2T1@0;
            F4 BY M2T2*1; M2T2@0;
            F5 BY M3T1*1; M3T1@0;
```

```
            F6 BY M3T2*1; M3T2@0;
            F1-F6@1;
            F1 WITH F2(P1);
            F1 WITH F3(P2);
            F1 WITH F4(P3);
            F1 WITH F5(P4);
            F1 WITH F6(P5);
            F2 WITH F3(P6);
            F2 WITH F4(P7);
            F2 WITH F5(P8);
            F2 WITH F6(P9);
            F3 WITH F4(P10);
            F3 WITH F5(P11);
            F3 WITH F6(P12);
            F4 WITH F5(P13);
            F4 WITH F6(P14);
            F5 WITH F6(P15);
MODEL CONSTRAINT:
            NEW(CV_DV);
            CV_DV = (P2+P7+P11+P14+P4+P9)/6
             - (P1+P3+P5+P6+P8+P10+P12+P13+P15)/9; ! MEAN DIFFERENCE
OUTPUT:     CINTERVAL;
```

Note. This command file differs from the preceding one only in its MODEL CONSTRAINT section. Specifically, after assigning a parametric symbol to each correlation coefficient, here the mean difference between convergent and discriminant validity coefficients of concern is defined in the quantity CV_DV (introduced first as a new parameter), and is subsequently interval estimated.