Do Two or More Multicomponent Instruments Measure the Same Construct? Testing Construct Congruence Using Latent Variable Modeling

Educational and Psychological
Measurement

2016, Vol. 76(5) 873–884

© The Author(s) 2015

Reprints and permissions:

sagepub.com/journalsPermissions.nav

DOI: 10.1177/0013164415604705

epm.sagepub.com

Tenko Raykov¹, George A. Marcoulides², and Bing Tong¹

Abstract

A latent variable modeling procedure is discussed that can be used to test if two or more homogeneous multicomponent instruments with distinct components are measuring the same underlying construct. The method is widely applicable in scale construction and development research and can also be of special interest in construct validation studies. The approach can be readily utilized in empirical settings with observed measure nonnormality and/or incomplete data sets. The procedure is based on testing model nesting restrictions, and it can be similarly employed to examine the collapsibility of latent variables evaluated by multidimensional measuring instruments. The outlined method is illustrated with two data examples.

Keywords

congruence, construct, latent variable modeling, multidimensionality, nested models, unidimensionality

¹Michigan State University, East Lansing, MI, USA

²University of California, Santa Barbara, CA, USA

Corresponding Author:

Tenko Raykov, Measurement and Quantitative Methods, Michigan State University, 443A Erickson Hall, East Lansing, MI 48824, USA.

Email: raykov@msu.edu

Educational, behavioral, social, biomedical, marketing, and business researchers are often confronted with settings where two or more multicomponent instruments are available to use for similar or presumably the same measuring purposes. Such concerns are particularly relevant in situations involving a single measurement occasion and instruments that are unidimensional as well as consist of distinct components. The same concerns are also of relevance in validation studies, specifically in empirical investigations in substantive areas where prior multiple lines of research have led to the development of several homogeneous instruments with different components, which purport to be measuring the same underlying latent variable.

As an approach to addressing these kinds of queries one might be inclined to consider at first evaluating the fit of a single-factor model that includes all instruments' components. This procedure would not necessarily yield a trustworthy result, however, owing to the fact that the typical application of such confirmatory factor analysis models in contemporary educational and psychological research is routinely focused mainly on overall goodness of fit indexes. These indexes are ill-equipped to respond to the actual question of relevance, which is whether two or more instruments measure the same construct. The reason is that the widely used overall fit indexes are merely omnibus measures of data-to-model discrepancy. For this reason, they may not be in a position to sense potentially serious and substantively meaningful "local" violations of unidimensionality in the single-factor model (unless, of course, these are strong enough to affect to a sufficient degree these overall indexes).

The present article discusses a widely applicable procedure that can be utilized to respond to these concerns, which are frequently of great importance in empirical research. The approach outlined below is developed within the latent variable modeling (LVM) methodology (see, e.g., Muthén, 2002) and is readily employed with popular LVM software in scale construction and development as well as construct validation studies. The method focuses on examining the plausibility of nesting restrictions whose violation need not necessarily affect sufficiently the overall goodness-of-fit indexes of the single-factor model in order to be sensed at their omnibus level. The procedure capitalizes on the relationships between models with increasing complexity as reflected in the number of latent variables evaluated by a given set of observed measures and is similarly applicable in cases of observed variable nonnormality as well as incomplete data sets. The method is illustrated using two examples that emphasize the necessity of testing more complex against less complex latent variable models, rather than only the latter as seems to be the widespread practice in similar circumstances in present-day empirical educational and behavioral research. The examples also demonstrate applications of the outlined LVM procedure when the goal instead is to respond in a more informed way to the question of main relevance: Do two or more homogeneous sets of distinct measures evaluate the same underlying latent construct?

Nesting Relationship Between Single-Factor and Multifactor Measurement Models

Background, Definitions, and Notation

The remainder of this article is based on an application of the LVM methodology for responding to the above-mentioned concerns. In particular, the following well-known factor analysis model will play an instrumental role:

$$y = \mu + \Lambda\eta + \varepsilon, \quad (1)$$

where y is the $p \times 1$ vector of distinct observed measures under consideration (overall set of instrument components, subscales, or subtests; $p > 3$ is an integer number), η is the $q \times 1$ vector of constructs underlying them ($q \geq 1$ is an integer number), Λ is the $p \times q$ factor loading matrix relating these measures to the constructs (see next for pertinent assumptions), and ε is the $p \times 1$ vector of unique factors comprising observed measure specificity and “pure” measurement error (see, e.g., Raykov & Marcoulides, 2011). In the rest of this discussion, it is similarly assumed that (i) all p measures in y are approximately continuous (see conclusion section for indicated method extension); (ii) y represents the overall collection of the components of q unidimensional (homogeneous, or congeneric) instruments of interest, and any observed measure in y loads on only one factor (i.e., Λ can be seen as a block-diagonal matrix with q blocks representing column vectors that correspond to the q constructs and are filled with free nonzero loadings, with all remaining elements of Λ being 0); (iii) the latent covariance matrix $\Phi = \text{Cov}(\eta)$ is positive definite, with units along its main diagonal and otherwise unrestricted (where $\text{Cov}(\cdot)$ denotes covariance matrix of the vector in parentheses); and (iv) Model (1) is plausible for a data set of concern in a given empirical setting where a researcher is interested in whether the set y evaluates a single construct (see also Jöreskog, 1971). For ease of reference, in the latter case the constructs in η will be referred to as “congruent” or “collapsible” (see Note 2 below). Last but not least, we assume that (v) the studied subjects are independent of each other, that is, there are no nesting effects of importance in a given sample or population under investigation (see conclusion section for additional discussion on this assumption).

The following definition of a model relationship of particular interest will be frequently used in the remainder of the article. We will say that two models are parameter-value nested (abbreviated “PV-nested”), if (a) they are based on/fitted to the same set of observed variables and (b) one of the models is obtained from the other model by setting (fixing) one or more parameters in the latter to be equal to prespecified numerical values.¹

The Relationship Between the Single-Factor and Multifactor Models

With the preceding definitions and assumptions (i) through (v), we are ready to introduce a main result in this article.

Proposition 1: Under the above assumptions, the single-factor model is PV-nested in model (1) for any value of $q \geq 2$.

To demonstrate this proposition, we can use the (mathematical) induction principle (see, e.g., Apostol, 2013). First, consider the case of $q = 2$. In it, due to the assumption of no cross-loadings (which is (ii) above), the two factors of concern load each on nonoverlapping subsets of the collection of observed variables that represent the set y of measures under consideration. Hence, when the correlation between these two factors (denoted ρ) is fixed to equal 1, then apart from an error term each measure in either subset is a deterministic linear function of any of the remaining measures in y , apart from its own error term. This is precisely the case, however, in the single-factor model, and only in that model (which for that reason could also be referred to as congeneric model; Jöreskog, 1971). Therefore, the single-factor model is PV-nested in the two-factor model, with the former being obtained from the latter by setting $\rho = 1$.² (A proof that as covariance structure models, the single-factor and two-factor models are nested according to the more general model nestedness definition, can be obtained from the authors on request; see Note 1.)

Assume next that Proposition 1 is true for a particular value of q , which is larger than 2, say for $q = k \geq 3$. That is, the single-factor model is by this assumption PV-nested in the k -factor model defined by Equation (1) (see also its subsequent discussion and in particular assumption (ii)). Therefore, for a certain set of say r parameters in the k -factor model ($r > 0$), it is true that if they are fixed correspondingly to prespecified values, the single-factor model is obtained from this k -factor model. Consider now the case of $q = k + 1$, that is, the $(k + 1)$ -factor model (1), where Λ is block-diagonal as assumed earlier (which is assumption (ii) above that is true here by virtue of the premise of Proposition 1). Select then any k from these $(k + 1)$ factors and their corresponding indicators. The collection of these indicators and their pertinent k factors represent obviously a k -factor model. Set (fix) in this model all those r parameters to their corresponding values, so that this activity according to the current assumption—viz., that Proposition 1 is true for $q = k$ —produces from this k -factor model the single-factor model. (Note that such setting/fixing for these r parameters is possible due to the single-factor model being PV-nested in the k -factor model following the last assumption.) Observe next that when this resulting single-factor model is extended by the remaining, $(k + 1)$ th factor and its pertinent indicators, one obtains a two-factor model. Then, in addition to these r parameter value restrictions producing the single-factor model from the k -factor model, add that of the correlation of the single factor in the former model with the remaining, $(k + 1)$ th factor being 1. Using the above reasoning for the case $q = 2$, which showed that the single-factor model is PV-nested in a two-factor model (1) (with (ii)), it follows now that the resulting model from all these $(r + 1)$ parameter restrictions is the single-factor model itself. Hence, the single-factor model is PV-nested in the $(k + 1)$ -factor model, and thus Proposition 1 is true also for $q = k + 1$. By the (mathematical) induction principle, therefore, Proposition 1 is true for any $q > 1$, as was to be shown in the first instance.

Utility of Proposition 1 for Validity Studies

Having now demonstrated Proposition 1 for any $q \geq 2$, we would like to emphasize its validity for the special case of $q = 2$ (i.e., two homogeneous instruments being studied for possibly measuring the same underlying construct). This case is of particular importance in construct validation studies, as alluded to earlier, and specifically when a new instrument is being developed with the aim of measuring the same construct like an earlier established instrument (see also conclusion section). Proposition 1 gives then a very useful result, viz., that the single-factor model is PV-nested in the two-factor model (with the above assumption (ii)), and hence also nested in the routinely used sense of the more general model nesting relationship routinely employed in LVM applications (see Note 1). Thereby, the PV-nesting restriction consists in the setting of perfect factor correlation. With this in mind, the latter constraint—viz., of unit factor correlation—is testable itself by testing the difference in fit in the two nested models involved, viz., the single-factor model and the two-factor model. This also represents a means of testing whether two homogeneous instruments with distinct components (e.g., a newly developed and an established instrument) in fact evaluate the same latent construct.³

When applying the outlined procedure in an empirical setting, we would like to emphasize that under the tested null hypothesis of unit correlation (in absolute value), the latter parameter is on the boundary of the pertinent parameter space. Hence, the p -value obtained with the used test statistic needs to be halved (or a corresponding p -value modification procedure is to be used in the general case with $q > 2$ instruments that is discussed next; see, e.g., Rabe-Hesketh & Skrondal, 2012; see also Notes 2 and 3).

General Construct Congruence and Collapsibility

In more general terms, Proposition 1 gives a researcher a means to test whether the latent constructs underlying any number q ($q > 1$) of unidimensional multicomponent instruments with distinct components, are identical up to linear transformations (see Note 3). This is also a particularly interesting substantive phenomenon, which was referred to above as “construct congruence” or “construct collapsibility.” Indeed, since the nesting restrictions are in fact the factor correlation identities to 1 (see Note 2 and the preceding demonstration of Proposition 1), it follows that testing for construct congruence or collapsibility is equivalent to testing whether the single-factor model has the same overall fit to an analyzed data set as the corresponding q -factor model for the entire set of observed measures (components) pertaining to the q instruments considered. As is well known, the last test can be carried out using the classical “chi-square difference test” in case of component normality (see, e.g., Bollen, 1989). When the latter assumption is violated, the corrected chi-square difference test can be used instead (excluding cases with piling of units of analysis at an extreme end for a component, or strong nesting effect within higher-order units; see, e.g., Raykov & Marcoulides, 2011; see also conclusion section, and the preceding subsection pointing out the need then for a p -value modification in this general case as well; Rabe-Hesketh & Skrondal, 2012).

Testing Construct Congruence

As a result of the preceding discussion in this article, we can now state that the last mentioned (corrected) chi-square test of the difference in fit of the single-factor and q -factor models represents a test of construct congruence/collapsibility for the q constructs underlying q given homogeneous measuring instruments with distinct components, for any number q of such instruments under consideration ($q \geq 2$). This test can be also directly carried out when incomplete data sets are available, including also auxiliary variables if the assumption of data missing at random (MAR) is suspected of being violated. (On the definition and choice of such variables, see, e.g., Enders, 2010; see also second example in the following section.)

We illustrate next the utility and applicability of the discussed LVM procedure for examining construct congruence using two numerical examples.

Illustration on Data

For the purpose of demonstrating the outlined method in this article, two examples are used—one with a simulated data set and the second with an empirical data set.

Lack of Construct Congruence

In the first example, we simulated a data set on $n = 600$ cases from $q = 2$ homogeneous instruments each with 7 and 5 components, respectively, which were generated according to the following model:

$$\begin{aligned}
 y_1 &= 3 + .6 \eta_1 + \varepsilon_1, \\
 y_2 &= 3 + .65 \eta_1 + \varepsilon_2, \\
 y_3 &= 3 + .7 \eta_1 + \varepsilon_3, \\
 y_4 &= 3 + .75 \eta_1 + \varepsilon_4, \\
 y_5 &= 3 + .8 \eta_1 + \varepsilon_5, \\
 y_6 &= 3 + .6 \eta_1 + \varepsilon_6, \\
 y_7 &= 3 + .65 \eta_1 + \varepsilon_7, \\
 y_8 &= 3 + .8 \eta_2 + \varepsilon_8, \\
 y_9 &= 3 + .8 \eta_2 + \varepsilon_9, \\
 y_{10} &= 3 + .8 \eta_2 + \varepsilon_{10}, \\
 y_{11} &= 3 + .8 \eta_2 + \varepsilon_{11}, \\
 y_{12} &= 3 + .8 \eta_2 + \varepsilon_{12},
 \end{aligned} \tag{2}$$

where η_1 and η_2 were standard normal variates with correlation $\rho = .75$, while the error terms were independent standard normal variables. (Further details on the data simulation process can be obtained from the authors on request.)

We first fit the single-factor model to all $p = 12$ measures (using the maximum likelihood method; see, e.g., Bollen, 1989). This yields the following goodness-of-fit indexes: chi-square = 140.089, degrees of freedom (df) = 54, root mean square error of approximation (RMSEA) = .052, with a 90% confidence interval (CI) of (.041, .062). These results can be considered indicative of a tenable model (see in particular the left end-point of the CI for the RMSEA; see also Hu & Bentler, 1999). We can therefore conclude that the findings suggest the single-factor model as plausible for the analyzed data set.

With this in mind, the presented example is a demonstration of a situation where the incorrect unidimensionality hypothesis for the collection of all $p = 12$ measures (across both instruments) fails to be rejected by the routinely used goodness of fit indexes in contemporary educational, behavioral, and social research, owing to the fact that they represent as mentioned overall measures of model-to-data discrepancy (see Equations (2) and the discussion following immediately). We also notice that by the data generation mechanism, either construct explains merely about half of the variance in the other, leaving thus the remaining almost half unexplained; that is, the two constructs are created as quite distinct from each other by the data simulation mechanism to begin with.⁴

Fitting next the two-factor model (1), with its constructs loading correspondingly only on the first 7 and on the last 5 of the 12 observed measures, similarly leads to the following tenable goodness-of-fit indexes: chi-square = 58.374, $df = 53$, RMSEA = .013 (0, .030). These results are of course consistent with the data generating mechanism utilizing two underlying latent variables (see Equations (2)).

To complete the application of the outlined LVM procedure for examining construct congruence, we conduct the likelihood ratio test of the null hypothesis $H_0: \rho = 1$, for the correlation ρ of the two factors in the last fitted model. To this end, as elaborated earlier, we need to carry out the “difference in chi-square test” (e.g., Bollen, 1989). Accordingly, the test statistic of the null hypothesis (H_0) is $140.089 - 58.374 = 81.715$, and hence is significant—that is, the associated p -value is less than .05—when compared to the pertinent chi-square distribution with 1 ($= 54 - 53$) degree of freedom. This is due to the fact that the cutoff of the chi-square distribution with $df = 1$ at the usual significance level $\alpha = .05$ is 3.84. We stress however that, as mentioned earlier in the article, under the null hypothesis being tested the fixed correlation parameter is at the boundary of the parameter space, and so the obtained p -value needs to be halved, making the relevant p -value in the end even smaller and of course still significant (see, e.g., Rabe-Hesketh & Skrondal, 2012). We thus reject the null hypothesis H_0 , which as indicated earlier is equivalent to that of construct congruence, that is, to the statement of $q = 1$ factor underlying the $p = 12$ measures under consideration. This is a correct conclusion, given the underlying data generation process (see Equations (2) and Note 3). This conclusion was however not arrived at with the

routinely utilized overall goodness-of-fit indexes commonly used in present-day educational, behavioral, and social research, as seen above; rather, the correct conclusion of the presence of two rather than a single construct behind these 12 measures was made possible by use of the outlined procedure in this article.

Construct Collapsibility

In the second example, we utilize data from a cognitive aging study by Baltes, Dittmann-Kohli, and Kliegl (1986), who were in part interested in examining older adults' performance on measures of the fluid intelligence subability Induction (Inductive reasoning; see, e.g., Horn, 1982). For illustration purposes, we use data from $n = 116$ elderly persons on two pairs of tests of inductive reasoning correspondingly with respect to letters and figures. Due to some nonnormality observed on the 4 measures, we employ in the remainder the robust maximum likelihood method of model fitting (see, e.g., Muthén & Muthén, 2014). In addition, since a considerable part of the studied elderly persons did not complete at least one of the tests, and that as usual the MAR assumption cannot be necessarily considered fulfilled for the resulting incomplete data set, we include as auxiliary variables two figural relations, culture-fair, and Raven's advanced progressive matrices tests that were also part of the test battery administered in the original investigation (see, e.g., Enders, 2010). Further details on the empirical study can be found in Baltes et al. (1986), and for additional discussion on the utility of incorporating auxiliary variables in missing data analyses with violations of MAR we refer readers to Graham (2003, 2009).

We begin by fitting the single-factor model to the 4 inductive reasoning tests of concern in this example, and find that its overall goodness-of-fit indexes can be considered tenable: chi-square = 1.789, $df = 2$, RMSEA = 0 (0, .178) (scale correction factor for robust maximum likelihood = 1.257; see, e.g., Muthén & Muthén, 2014). As a next step, we fit the two-factor model (1), with correlated factors that load each only on the pertinent two of the inductive reasoning tests, and find that its goodness-of-fit indexes can also be viewed as tenable: chi-square = .938, $df = 1$, RMSEA = 0 (0, .242) (scale correction factor for robust maximum likelihood = 1.071). To complete now the application of the procedure discussed in this article, we apply the "corrected difference in chi-square test"—which we need here due to using the robust maximum likelihood method to fit the models—for testing the null hypothesis of the two factors being collapsible, that is, their correlation being 1. To this end, we can use the R-function provided in the appendix (see www.statmodel.com for pertinent test statistic formula that is implemented in the function), which renders the following results: corrected chi-square difference = 0.590, $df = 1$, p -value = .442. To complete the application of the outlined procedure in this article, we emphasize that, as mentioned earlier, under the null hypothesis being tested the fixed correlation parameter is at the boundary of the parameter space, and so the last reported p -value needs to be halved, making the relevant p -value equal to .221, which is still nonsignificant (see, e.g., Rabe-Hesketh & Skrondal, 2012). These findings suggest that the null

hypothesis of construct collapsibility (congruence) is tenable, that is, one may consider as plausible the conjecture that a single factor—in this example, the construct of inductive reasoning ability (as it may be referred to)—rather than two distinct ones as in the first fitted model in this example, underlies the performance on the considered 4 induction tests in the studied population of elderly persons (see Note 3).

Conclusion

Educational, behavioral, and social scholars involved in scale construction and development, and in particular in validation studies, are often confronted with the need to evaluate whether two unidimensional instruments may be measuring the same underlying construct. In more general terms, it may well be relevant in an empirical study, in particular when concerned with validity of measurement, to examine whether several constructs underlying a multidimensional instrument may in fact be collapsible into a single latent variable explaining performance on its components.

While fitting the popular single-factor model to the entire set of instrument components addresses part of these concerns, it does not in general respond in a satisfactory manner to the main question of relevance in such settings, viz., whether two or more homogeneous instruments measure the same underlying construct (see Note 3; see also the first example in the illustration section). The present article discussed a widely applicable LVM procedure for accomplishing this aim, which is based on testing PV-nesting restrictions relating the single-factor with respective multiple-factor models. The outlined method is readily utilized with popular statistical software, and is directly applicable with nonnormal data as well as with incomplete data sets (assuming availability of useful and effective auxiliary variables; see, e.g., Enders, 2010; Graham, 2003, 2009).

The approach to examining construct congruence outlined in this article has several limitations that are worth pointing out here. First, the discussed procedure is based on the assumption of large samples, since its method of parameter estimation and model testing is based on maximum likelihood (robust maximum likelihood), which itself is grounded in an asymptotic statistical theory. Second, as pointed out earlier, the method assumes that all instruments' components are approximately continuous. Pending future research on the relevance of the asymptotic theory at particular finite samples, as well as the actual sensitivity of the robust maximum likelihood method to violations of normality in the form of indicator discreteness, it may be recommended to use meanwhile this method also with observed measures that have at least 5 to 7 possible values (and large samples with no piling of cases at scale end or considerable nesting effects; see, e.g., Raykov & Marcoulides, 2011). Third, with (approximately) continuous measures that are not more than mildly nonnormal, use of robust maximum likelihood may also be recommendable with relatively minor violations of the assumption of no nesting effects that underlies the article. Fourth, while this is not a limitation per se of the outlined method, it is important to keep in mind that under the tested null hypothesis of unit correlation(s) (in absolute value), the

latter parameter(s) is on the boundary of the pertinent parameter space, and hence the p -value obtained with the used test statistic needs to be halved (or correspondingly modified in the general case; see, e.g., Rabe-Hesketh & Skrondal, 2012; see also Notes 2 and 3). Last but not least, with respect to the assumption of (approximately) continuous observed measures, we do not see any obstacles in principle to extend directly the method of the article to the case of binary or highly discrete observed variables (with up to 5-7 possible values, large samples, and no piling of cases or notable nesting effects). This extension will utilize the same procedure as outlined earlier in the article, but on the assumed underlying normal variables—as usual in LVM with discrete observed measures (see, e.g., Muthén, 2002). That method extension is however beyond the confines of the present article.

In conclusion, this article outlined a readily and widely applicable LVM procedure for examining construct congruence (collapsibility), which should be of particular interest to educational, behavioral, social, biomedical, and marketing scientists involved in scale construction and development and especially in validity studies. The outlined approach enlarges the arsenal of measurement methods available to these scientists by a procedure that allows them to respond in a more informed way to the often important question whether two or more homogeneous instruments are measuring the same underlying latent construct.

Appendix

R-Function for Conducting the Corrected Chi-Square Difference Test on Nested Models

```
corr.chisq.diff.test = function(d0, d1, c0, c1, t0, t1){
  t = (t0 - t1)*(d0-d1)/(d0*c0 - d1*c1)
  p=1-pchisq(t,d0-d1) # d - df's, c - scaling correction factors,
  # t - (corrected) chi-square values; all output by Mplus
  c(t,d0-d1, p) # 0 - index of nested model, 1 - index of full model
} # see next note related to final p-value determination.
```

Note. Once pasting this function at the R prompt, call it by providing the two models' chi-square values, degrees of freedom, and scaling correction factors, in the order listed in the first line of the above code (comments preceded by “#” can be dropped thereby; for a nontechnical introduction to the freely available software R, see, e.g., Raykov & Marcoulides, 2012; see also main text and Rabe-Hesketh & Skrondal, 2012, for a p -value correction needed in empirical applications of this R-function, due to parameter space boundary value(s) under the tested null hypothesis).

Acknowledgments

Thanks are due to D. Gunzler and J. Dalton for valuable discussions on structural equation modeling applications.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Notes

1. As implied from this definition, two models that are PV-nested are also nested according to the more general definition of nested models that is routinely utilized in LVM applications. (For the latter, general definition to apply, as is well known one of the models in a nested pair has to result from the other model by imposing appropriate parameter constraints.) We use in the remainder of this article the definition of PV-nestedness stated in the main text, in order to emphasize that unlike this definition that routinely used and more general definition includes also linear and nonlinear parameter restrictions as producing a nested model from a given one; in difference to the latter definition, the PV-nestedness definition is only concerned with restrictions that set/fix particular parameters as equal to prespecified constants. This type of numerical value setting restrictions are the only ones of relevance and interest in the present article.
2. None of the following developments will be affected if $\rho = -1$ were to hold. Since most of the times in empirical educational and behavioral research observed measures are positively correlated (or can be correspondingly recoded to be so), without loss of generality we limit the discussion in this article to the case of positive factor correlation(s).
3. We should like to point out that if this nested model test turns out not to be significant in an empirical study, all one can say is that there is not sufficient evidence in the analyzed data to warrant rejection of the null hypothesis stipulating the two constructs as being perfect linear combinations of each other (which is equivalent to the null hypothesis of their correlation being 1; see also Note 2). In this case of retaining that null hypothesis, we talk in the present article of “construct congruence” or “construct collapsibility.” We stress, however, that the latter is not necessarily sufficient to declare the two constructs as strictly identical, but merely as being linear transformations of each other. How this finding could be further interpreted in a particular empirical study is deferred to substantive experts in the pertinent subject-matter domain.
4. The fit of the single-factor model to the first 7 observed variables only (i.e., considered separately from the last 5 variables) was found to be tenable as well: chi-square = 21.028, $df = 14$, RMSEA = .029 (0, .053). Similarly, the fit of the single-factor model for the last 5 variables only (separately from the first 7) was tenable, too: chi-square = 8.370, $df = 5$, RMSEA = .034 (0, .072).

References

- Apostol, T. M. (2013). *Calculus*. New York, NY: Wiley.
- Baltes, P. B., Dittmann-Kohli, F., & Kliegl, R. (1986). Reserve capacity of the elderly in aging-sensitive tasks of fluid intelligence: Replication and extension. *Psychology and Aging, 1*, 172-177.

- Bollen, K. A. (1989). *Structural equations with latent variables*. New York, NY: Wiley.
- Enders, C. K. (2010). *Applied missing data analysis*. New York, NY: Guilford.
- Graham, J. W. (2003). Adding missing-data relevant variables to FIML-based structural equation models. *Structural Equation Modeling*, 10, 80-100.
- Graham, J. W. (2009). Missing data analysis: Making it work in the real work. *Annual Review of Psychology*, 60, 549-575.
- Horn, J. L. (1982). The aging of human abilities. In B. B. Wolman (Ed.), *Handbook of developmental psychology* (pp. 847-870). Englewood Cliffs, NJ: Prentice Hall.
- Hu, L.-T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1-55.
- Jöreskog, K. G. (1971). Statistical analysis of sets of congeneric tests. *Psychometrika*, 36, 109-133.
- Muthén, B. O. (2002). Beyond SEM: General latent variable modeling. *Behaviormetrika*, 29, 81-117.
- Muthén, L. K., & Muthén, B. O. (2014). *Mplus user's guide*. Los Angeles, CA: Muthén & Muthén.
- Rabe-Hesketh, S., & Skrondal, A. (2012). *Multilevel and longitudinal modeling using Stata*. College Station, TX: Stata Press.
- Raykov, T., & Marcoulides, G. A. (2011). *Introduction to psychometric theory*. New York, NY: Taylor & Francis.
- Raykov, T., & Marcoulides, G. A. (2012). *Basic statistics: An introduction with R*. New York, NY: Rowman & Littlefield.