Article

Educational and Psychological Measurement
2018, Vol. 78(5) 905–917

DOI: 10.1177/0013164417698017
journals.sagepub.com/home/epm
SAGE

Studying Latent Criterion Validity for Complex Structure Measuring Instruments Using Latent Variable Modeling

Tenko Raykov[1], Natalja Menold[2], and George A. Marcoulides[3]

Abstract

Validity coefficients for multicomponent measuring instruments are known to be affected by measurement error that attenuates them, affects associated standard errors, and influences results of statistical tests with respect to population parameter values. To account for measurement error, a latent variable modeling approach is discussed that allows point and interval estimation of the relationship of an underlying latent factor to a criterion variable in a setting that is more general than the commonly considered homogeneous psychometric test case. The method is particularly helpful in validity studies for scales with a second-order factorial structure, by allowing evaluation of the relationship between the second-order factor and a criterion variable. The procedure is similarly useful in studies of discriminant, convergent, concurrent, and predictive validity of measuring instruments with complex latent structure, and is readily applicable when measuring interrelated traits that share a common variance source. The outlined approach is illustrated using data from an authoritarianism study.

Keywords

correlation, criterion validity, factor analysis, latent variable modeling, measurement error, second-order factor structure, validity

[1]Michigan State University, East Lansing, MI, USA
[2]Leibniz Institute for the Social Sciences, Mannheim, Germany
[3]University 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

Validity is an essential requirement of measurement in the behavioral, educational, and social sciences (e.g., McDonald, 1999). A main type of validity is criterion validity that is subsumed under the comprehensive construct validity concept (Messick, 1995). A commonly used index of criterion validity is the correlation coefficient between a test or scale score and a criterion variable, especially when both can be considered or treated as continuous (e.g., Raykov & Marcoulides, 2011; see also Crocker & Algina, 2006, for the discrete case). As is well known, however, measurement error usually attenuates this coefficient (cf. Raykov, Marcoulides, & Patelis, 2015, for a more general treatment). Accounting for this error both in the scale score and criterion variable is therefore an essential step toward obtaining unbiased estimates of criterion validity coefficients.

Multicomponent instrument construction is a complicated multistage process (e.g., Raykov, 2012). Despite well-informed efforts aimed at attaining unidimensionality of the resulting instrument—such as a test, scale, inventory, or self-report (referred to also as ‘‘test’’ or ‘‘scale’’ below)—important concerns about validity and construct underrepresentation tend to contribute to its more complex latent structure in empirical research. Oftentimes this test may be tapping into more than one interrelated constructs that load on a second-order factor. For example, a mathematics ability test could consist of a part evaluating algebra ability, another assessing geometry ability, a third measuring trigonometry ability, and a fourth concerned with problems assessing abstract thinking ability, with all four abilities loading on a second-order factor representative of the targeted mathematics ability. In such situations, it may be difficult to argue for a wider use of the overall scale score (the unweighted or weighted sum of the individual instrument components), owing to the fact that the latter is not unidimensional (see also below). For this reason, it would similarly be problematic to interpret the correlation of that scale score with a criterion variable of interest as a straightforward index of criterion validity.

The present article addresses these concerns using a latent variable modeling (LVM; e.g., Muthén, 2002) approach. An LVM procedure is discussed below that allows accounting for (a) the measurement error in the overall scale or test score as well as in the criterion variable(s), and (b) the second-order factorial structure of an instrument under consideration (see also Note 1). The method permits one to point and interval estimate the correlation between the second-order factor and a criterion variable. With that feature of the approach, this correlation may be considered a latent criterion validity coefficient that one could argue represents an appropriate validity index in such complex latent structure settings. The procedure can be viewed as a generalization of the method of discriminant and convergent validity evaluation discussed in Raykov and Marcoulides (2011, chap. 8), which assumes unidimensionality of a measuring instrument under consideration, to the case of lack of unidimensionality that is characterized by a second-order factorial structure. The discussed method is illustrated on data from a study involving the measurement of a multidimensional concept of authoritarianism (Beierlein, Asbrock, Kauff, & Schmidt, 2014; Duckitt & Bizumic, 2013).

Background, Notation, and Assumptions

For the aims of the present article, we assume that a set of (approximately) continuous measures are given that are denoted $Y_1, Y_2, \ldots, Y_p$ ($p > 1$) and represent the components of a measuring instrument whose criterion validity is of interest to evaluate with respect to a prespecified variable. For this test or scale, we posit the following second-order factor structure

$$y = \underline{\mu} + \Lambda\underline{\eta} + \underline{\varepsilon} \tag{1}$$

$$\underline{\eta} = \Gamma\xi + \underline{\delta} \tag{2}$$

where $\underline{y} = (Y_1, Y_2, \ldots, Y_p)'$ is the $p \times 1$ vector of test components, $\underline{\mu}$ is the $p \times 1$ vector of associated intercepts, $\underline{\eta}$ is the $q \times 1$ vector of first-order common factors assumed with zero mean and a positive definite covariance matrix ($q > 0$), ξ is a second-order factor with zero mean and positive variance, Λ is a $p \times q$ matrix of first-order factor loadings while Γ is the $q \times 1$ second-order factor loading matrix, $\underline{\varepsilon}$ is the $p \times 1$ vector of unique factors with zero means and a diagonal positive definite covariance matrix that are uncorrelated with ξ and $\underline{\eta}$, and $\underline{\delta}$ is a $p \times 1$vector of residual terms with a positive definite covariance matrix that are uncorrelated with ξ, $\underline{\eta}$, and $\underline{\varepsilon}$ (e.g., Harman, 1976; priming denotes transposition and underlining vector in this article). The model defined in Equations (1) and (2), when appropriately extended to include a criterion variable(s) and identified as well as plausible for a studied population, will play an instrumental role in the remainder of this article. Throughout the article, we also assume that the population under investigation consists of independent cases (persons; see Conclusion section for a possible extension).

A Latent Variable Modeling Procedure for Evaluating Latent Criterion Validity of Measuring Instruments With Second-Order Factorial Structure

Equation (1) shows that it would not be appropriate, strictly speaking, to view as unidimensional the instrument comprising the components $\underline{y}$, despite its feature of evaluating the second-order factor ξ. Indeed, substituting Equation (2) into (1) we obtain

$$\begin{aligned} y &= \underline{\mu} + \Lambda(\Gamma\xi + \underline{\delta}) + \underline{\varepsilon} \\ &= \underline{\mu} + \Lambda\Gamma\xi + \Lambda\underline{\delta} + \underline{\varepsilon} \end{aligned} \tag{3}$$

From Equation (3), after straightforward algebra (e.g., on the individual observed variable equations), one notices that all observed variables loading say on η_j in Equation (1) share the (common) factors ξ and δ_j ($j = 1, \ldots, q$); hence, the instrument consisting of the measures $\underline{y}$ is not homogeneous. (Notice that δ_j cannot be really considered part of a residual term in model (3), since δ_j is in fact part of the jth first-order common factor η_j and not of its unique factor ε_j; $j = 1, \ldots, q$.) For this reason, correlating the overall test or scale sum score

$$X = y_1 + y_2 + \cdots + y_p \tag{4}$$

or the weighted sum

$$W = w_1 y_1 + w_2 y_2 + \cdots + w_p y_p \tag{5}$$

using weights w_j ($j = 1, \ldots, p$), with a criterion variable cannot be meaningfully treated as a criterion validity index associated with this multicomponent measuring instrument.

A Latent Criterion Validity Coefficient

Equations (3) through (5) reveal further that for any given value of the second-order factor ξ, the average sum scores X and W represent deterministic linear functions of ξ. Hence, one can treat each of these sum scores as effectively measuring that higher order factor ξ.

In order to be in a position to account for measurement error in a criterion variable of interest, we assume it in the rest of this discussion as a latent variable (construct) ζ with positive variance that is evaluated by m indicators, Z_1 through Z_m ($m > 1$). The remainder also assumes, as indicated earlier, that the overall model resulting when Equations (1) and (2) are augmented by $Z_1, \ldots, Z_m$ and the pertinent measurement model for ζ (see Equation 7 for its formal definition) is identified as well as plausible for a studied population.[1]

For this setting, which is frequently of relevance in empirical social and behavioral research, the present article proposes to consider the correlation

$$v = Corr(\zeta, \xi) \tag{6}$$

where $Corr(\cdot,\cdot)$ denotes correlation, as an appropriate criterion validity coefficient associated with the multi-component instrument comprising the measures $\underline{y}$, and refer to the quantity in Equation (6) then as a latent criterion validity coefficient (LCVC).

Point and Interval Estimation of Latent Criterion Validity

In order to estimate and obtain a confidence interval of the LCVC, one can make use of LVM. To this end, we consider the model

$$\begin{aligned} y &= \underline{\mu} + \Lambda\underline{\eta} + \underline{\varepsilon} \\ \underline{\eta} &= \Gamma\xi + \underline{\delta} \\ \underline{Z} &= \underline{\mu}_z + \Lambda_z\zeta + \omega \end{aligned} \tag{7}$$

which extends the model defined in Equations (1) and (2) (with its assumptions mentioned earlier) by the measurement model for the criterion indicators $\underline{Z}$, where $\underline{Z} = (Z_1, Z_2, \ldots, Z_m)'$ is the $m \times 1$ vector of indicators for the latent criterion ζ, $\underline{\mu}_Z$ is the $m \times 1$ vector of their intercepts, Λ_Z is the $m \times 1$ matrix of their loadings on ζ, and $\underline{\omega}$

is the associated $m \times 1$ vector of unique factors with zero means and a positive definite covariance matrix that are assumed uncorrelated with ζ, ξ, and $\underline{\varepsilon}$.

With model (7) in mind, the LCVC in Equation (6) is readily seen as a nonlinear function of appropriate model parameters, namely, as the ratio of the latent covariance of the second order factor and latent criterion to the product of their square rooted variances (cf. Crocker & Algina, 2006). Hence, point and interval estimation of the LCVC proposed in this article becomes possible when model (7) is fitted to data and found plausible. This fitting is feasible using for instance the popular LVM software Mplus (Muthén & Muthén, 2017), and a point estimate of the LCVC (6) is rendered thereby as a routine product of this process. An application subsequently of the monotone transformation-based approach to confidence interval estimation of the correlation coefficient in Equation (6), as described for instance in Raykov and Marcoulides (2011), furnishes a confidence interval (CI) for the LCVC of interest here. (This CI construction approach is readily applied using the R-function ''ci.pc'' in the last cited source, which is also presented in the appendix for completeness of the present discussion.) The Mplus source code accomplishing the LCVC evaluation is also provided in the appendix, where it is applied on the empirical data used in the next section to demonstrate the outlined criterion validity evaluation procedure.

Illustration on Empirical Data

For the aims of this section, we utilize data from a study of n = 163 members of an online panel representing a sample of German adults (internet users), which was concerned with examining right-wing authoritarianism (Duckitt & Bizumic, 2013). Right-wing authoritarianism has been defined as a multidimensional concept covering three main traits that were identified by Adorno, Frenkel-Brunswik, Levinson, and Sanford (1950) as follows: (i) conformity to the governmental structures and state authorities (submissiveness), (ii) authoritarian aggression, and (iii) rigid support for traditions and established norms (conventionalism). According to Altemeyer (1981), the combination of these three traits leads to the concept of right-wing authoritarianism.

In the presently used empirical study, authoritarianism as a second-order factor was measured by the Short Scale of Authoritarianism (referred to as KSA-3; Beierlein et al., 2014). The scale consists of 9 indicators (items) and covers the above dimensions (i) through (iii), with three items per dimension, which are correspondingly referred to as Aggression, Submission and Conventionalism in the rest of the section (see Table 1 for specifics regarding these items per dimension). The nine indicators of authoritarianism were verbally formulated in such a way that high scores (or agreement with their statements) were associated with a higher degree of authoritarianism. As a latent criterion variable, we utilize gender-role attitude that was evaluated by three items on ''consequences for parenting'', and refer to it as Parenting in the remainder (see also Note 1). These items constitute a subscale of

Table 1. KSA-3 (*Die Kurzskala Autoritarismus*) and Parenting Items (Authors' Translation From German).

Authoritarianism items (source: Beierlein et al., 2014; response categories range from *do not agree at all* to *fully agree*, on a 5-point rating scale)

1. Use of the strongest means possible is justified, in order to neurtralize unproductive elements of society.
2. Troublemakers should be made aware that they are not welcome in our society.
3. Societal regulations should be enforced without reservation.

Submission items:

4. What we need to live in safety is a strong, determined leader.
5. People should delegate to the authorities decisions that are important for the society.
6. We should be grateful to leaders that tell us exactly what we need to do.

Conventionalism items:

7. Traditions should be cultivated, respected, and upheld.
8. Behaviors proven appropraite in society should not be questioned.
9. It is always best to do things in the manner accepted by society.

Parenting Items (source: GESIS, 2016; response categories range from *does not apply at all to me* to *applies fully to me*, on a 7-point rating scale):

1. An employed mother can have as loving and close relationship with her children as an unemployed mother can.
2. A young child would suffer, if her/his mother is working.
3. It is better for a child if his/her mother is employed and does not only concentrate on household work.

a psychometric battery evaluating gender-role attitudes that was included in the German General Social Survey (GGSS) study (cf. Braun, 2014). Two indicators (items) of Parenting were formulated in such a way that agreement (higher scores) expressed a liberal opinion and support for women labor force participation, while one item was reversely formulated (in keeping with modern approaches to self-reporting, the wording of both the positive and negative items was always positive; Ferrando & Lorenzo-Seva, 2010). The text of all 12 items from the KSA-3 (*Die Kurzskala Autoritarismus*) scale and the Parenting subscale, is provided in Table 1 (with an approximate translation from their original German version).

To illustrate the validity evaluation method outlined in the preceding sections of the present paper, we fit to this study data model (7) that effectively underlies this article, with $p = 9$ and $q = m = 3$ (see the appendix for the needed Mplus source code and notes to it). This model includes a total of 5 latent constructs—the above three first-order factors of Aggression, Submission, and Conventionalism with three indicators each, their second-order factor Authoritarianism, and the criterion construct of Parenting. The indicators of Authoritarianism were evaluated each using a 5-point numeric fully verbalized rating scale and the indicators on parenting were evaluated using a 7-point numeric rating scale (see also Table 1). The 12 indicators are considered for the illustration purposes of this section as approximately continuous measures on which the robust maximum likelihood method of model testing and parameter estimation is applied (cf. DiStefano, 2002; see also Raykov & Marcoulides, 2011). To deal with a notable proportion of missing data in the construct indicators and counteract possible violations of the missing at random (MAR) assumption underlying this method, we also include as an auxiliary variable the score from the so-called Left-Right Self-Placement scale (LRSP; e.g., Enders, 2010). LRSP refers to the two familiar political orientations (with ''left'' being associated here with liberal positions, described for instance by loyalty and acceptance of other groups' positions and low levels of concerns with respect to social order and obligations; and ''right'' being associated with conservative political opinions, described for example by strong in-group orientations and high respect for social order, established norms and hierarchies; e.g., Crawford & Pilanski, 2014; Haidt, Graham, & Joseph, 2009). Since both right-wing authoritarianism and gender-role attitudes are conceptually incorporated in the left-right distinction (Knight, 1993), and several studies have found notable relationship between them (e.g., Banaszak & Plutzer, 1993; Leone, Desimoni, & Chirumbolo, 2014), it was decided to use the LRSP scale score as an auxiliary variable (see also Enders, 2010).

The described model was found to be associated with the following tenable fit indices: chi-square (χ^2) = 64.419, degrees of freedom (*df*) = 50, *p*-value (*p*) = .083, and root mean square error of approximation (RMSEA) = .042 with a 90% confidence interval [0, 0.069]. The resulting parameter estimates in it are presented in Table 2.

In this plausible model, of particular interest is the LCVC estimate, which as seen from Table 2 results as −0.343, with a standard error of 0.123. Using the aforementioned monotone transformation-based approach to CI construction (see also the appendix), we obtain a 95% CI for the LCVC of (−0.558, −0.084). This relatively wide CI is not unexpected given the sample size that cannot be considered really large in the empirical study used. At least as importantly, the CI indicates a significant but weak to moderate (linear) relationship between Authoritarianism, as measured by the employed nine-item scale, on one hand, and the construct of Parenting as a criterion (latent) variable on the other hand. This correlation indicates a marked tendency of persons above average on Authoritarianism to be among those with scores on Parenting below its mean. In addition, as validity related coefficient, this correlation indicates also a considerable and expected discriminant validity of the used Authoritarianism scale with respect to the Parenting construct.

Conclusion

This article was concerned with an LVM procedure for point and interval estimation of a proposed latent criterion validity coefficient for a multicomponent measuring instrument with a latent structure that is more complex than that of unidimensionality. The discussed approach is useful in empirical situations with second-order factorial structure of psychometric tests or scales under consideration, when a researcher is also interested in accounting for measurement error in criterion variables of concern.

Table 2. Parameter Estimates, Standard Errors, *t* Values, and Two-Tailed *p* Values Associated With Fitted Model (Software Output Format).

Parameter	Estimate	S.E.	*t*-value	*p*-value
AGGR BY				
AGGR1	1.000	—	—	—
AGGR2	1.113	0.118	9.425	0.000
AGGR3	0.972	0.130	7.475	0.000
UW BY				
UW1	1.000	—	—	—
UW2	0.808	0.204	3.970	0.000
UW3	0.846	0.173	4.899	0.000
TRAD BY				
T1	1.000	—		—
T2	1.972	0.386	5.103	0.000
T3	1.189	0.231	5.156	0.000
PARENTNG BY				
RF11	1.000	—		—
RF12	−0.859	0.204	−4.211	0.000
RF13	0.727	0.167	4.345	0.000
A BY				
AGGR	1.000	—		—
UW	0.838	0.216	3.884	0.000
TRAD	0.487	0.176	2.762	0.006
A WITH				
PARENTNG	−0.259	0.108	−2.393	0.017
Intercepts				
AGGR1	2.671	0.085	31.592	0.000
AGGR2	3.037	0.088	34.548	0.000
AGGR3	2.598	0.088	29.577	0.000
UW1	3.042	0.079	38.485	0.000
UW2	1.909	0.066	29.108	0.000
UW3	1.786	0.069	25.993	0.000
T1	3.220	0.077	41.992	0.000
T2	2.323	0.080	29.062	0.000
T3	1.913	0.062	30.771	0.000
RF11	5.701	0.122	46.846	0.000
RF12	3.421	0.144	23.835	0.000
RF13	4.165	0.129	32.249	0.000
Variances				
A	0.354	0.133	2.665	0.008
PARENTNG	1.614	0.428	3.773	0.000
Residual Variances				
AGGR1	0.477	0.083	5.737	0.000
AGGR2	0.405	0.113	3.579	0.000
AGGR3	0.601	0.116	5.180	0.000
UW1	0.490	0.102	4.796	0.000
UW2	0.358	0.098	3.657	0.000
UW3	0.389	0.087	4.455	0.000

(continued)

Table 2. (continued)

Parameter	Estimate	S.E.	*t*-value	*p*-value
T1	0.737	0.090	8.182	0.000
T2	0.167	0.107	1.570	0.116
T3	0.307	0.057	5.372	0.000
RF11	0.815	0.370	2.206	0.027
RF12	2.188	0.376	5.817	0.000
RF13	1.883	0.306	6.146	0.000
AGGR	0.341	0.127	2.680	0.007
UW	0.282	0.083	3.422	0.001
TRAD	0.143	0.059	2.413	0.016
New/Additional Parameters				
LCV	−0.343	0.123	−2.795	0.005

Note. S.E. = standard error; AGGR# = indicator of Aggression construct; UW# = indicator of Submission construct; T# = indicators of Conventionalism construct; RF# = indicator of Parenting construct; A = Authoritarianism construct; LCV = Latent Construct Validity (LCVC in the text). The negative sign of RF12 is due to reversed scoring (see Table 1), and the negative sign of the LCV estimate indicates a tendency of high authoritarianism to be associated with lower support of labor force participation of women (see also main text).

The method is applicable when the criterion variables are uncorrelated with the error terms in the individual components of a given test or scale (a testable condition that can be examined using, e.g., the approach in Raykov, Marcoulides, Gabler, & Lee, 2017). The outlined procedure may be of particular utility in studies of criterion, discriminant, convergent, concurrent, or predictive validity of relatively long but internally consistent tests that are, however, not homogeneous.

Several limitations of the discussed approach are worthwhile pointing out here. As indicated earlier, the procedure assumes (approximately) continuous individual scale components and criterion variables. In case of indicator normality, as is well known the use of maximum likelihood (ML) estimation is appropriate and yields ML estimates of the latent criterion validity coefficients of interest (e.g., Bollen, 1989). With up to mild deviations from normality, which do not result from piling at scale end for an individual component(s), it may well be recommendable to use instead the robust ML method (MLR; Muthén & Muthén, 2017), possibly also with components having as few as five to seven response options (e.g., DiStefano, 2002; see also the appendix). Further research on the robustness of the MLR method is needed, however, in such situations. With fairly large samples, weighted least squares (WLS) estimation is also available with nonnormal continuous instrument components (e.g., Bollen, 1989). Relatedly, the outlined validity evaluation procedure is best used with large samples, owing to the fact that its application rests on ML, robust ML, or WLS estimation, with all of them grounded in asymptotic statistical theory (e.g., Muthén, 2002). Future research is needed also here, which may contribute to the development of possible guidelines for determining sample sizes at which one could rely on that large-sample theory.

As a third limitation, we assumed throughout that observations (studied persons) were independent, that is, not clustered or nested within (higher order) Level-2 units, such as schools, clinicians, interviewers, physicians, neighborhoods, cities, and so on. It may be hypothesized that the robust ML estimation method may also have some robustness to limited violations of this classical independence assumption, especially when the degree of nonnormality is not pronounced. To our knowledge, however, there is not sufficient research in this area that could help find out the extent and conditions under which one may trust such a potential recommendation.

Last but not least, the plausibility and identification of model (7) when used in applications of the procedure of this paper is essential, as indicated earlier (see also Note 1). When either of these two conditions is not satisfied, the discussed method cannot be generally recommended as it may yield misleading parameter estimates, standard errors, and statistical test results with regard to criterion validity of studied tests or scales. Lack of identification of the overall model may be expected with an insufficient number of indicators for any of the first-order factors, and may be resolved by adding appropriate parameter constraints that reflect substantively plausible parameter relationships in studied populations (e.g., Raykov & Marcoulides, 2006).

In conclusion, this article offers to empirical educational, behavioral, and social scientists a widely applicable means for point and interval estimation of criterion validity of multicomponent measuring instruments with second-order factorial structure, which permits also accounting for measurement error in associated overall sum scores (whether weighted or not) as well as in used criterion variables.

Appendix

Mplus Source Code for Evaluating Latent Criterion Validity

```
TITLE:      MPLUS SOURCE CODE FOR EVALUATING LATENT CRITERION VALIDITY.
            (ANNOTATING COMMENTS ADDED AFTER EXCLAMATION MARK.)
DATA:       FILE = <NAME OF RAW DATA FILE > ;
VARIABLE:   NAMES = AGGR1-AGGR3 UW1-UW3 T1-T3 LRSE RF11RF12 RF13 RF21
            RF22 RF23;
            MISSING = ALL(-999);
            AUXILIARY = (M) LRSE; ! TO COUNTERACT POSSIBLE VIOLATIONS OF MAR.
ANALYSIS:   ESTIMATOR = MLR;
MODEL:      AGGR BY AGGR1-AGGR3;
            UW BY UW1-UW3;
            TRAD BY T1-T3;
            A BY AGGR UW TRAD;
            A(V_A); ! PARAMETER TO BE USED IN MODEL CONSTRAINT SECTION.
            PARENTNG BY RF11-RF13; ! LATENT CRITERION VARIABLE/CONSTRUCT.
            PARENTNG(V_P); ! PARAMETER TO BE USED BELOW.
            A WITH PARENTNG (COV_AP); ! PARAMETER TO BE USED BELOW.
```

```
MODEL CONSTRAINT:
                NEW(LCVC); ! CREATING A PLACE HOLDER FOR LCVC.
                LCVC = COV_AP/SQRT(V_A*V_P); ! SEE EQUATION (6).
OUTPUT:         CINTERVAL;
```

Note 1. Variable notation identical to one used in Table 2 (see note to Table 2). For an introduction to the Mplus syntax, see, for example, Raykov and Marcoulides (2006).

Note 2. To obtain say a 95% confidence interval for the LCVC, use the following R-function that is presented here only for completeness purposes (see R-function ''ci.pc'' in Raykov & Marcoulides, 2011, chap. 8; for another confidence level, use pertinent cutoff value rather than the constant 1.96 below):

```
ci.lcvc <- function(c, se){
 # R-function for interval estimation of the LCVC z=.5*log((1+c)/(1-c))
 sez = se/((1-c^2))
 ci_z_lo = z-1.96*sez
 ci_z_up = z+1.96*sez
 ci_lo = (exp(2*ci_z_lo)-1)/(exp(2*ci_z_lo)+1)
 ci_up = (exp(2*ci_z_up)-1)/(exp(2*ci_z_up)+1)
 ci = c(ci_lo, ci_up)
ci
}
```

Acknowledgments

This research was in part conducted while T. Raykov was visiting the Leibniz Institute for the Social Sciences, Mannheim, Germany. Thanks are due to B. Rammstedt for helpful support.

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.

Note

1. The assumption of the criterion being a latent variable with multiple indicators is not essential for the method of this paper. In fact, this method is applicable also with a single criterion measure assumed to be error-free, after a minor modification amounting to representing the latter as identical to a latent dummy variable associated with a zero variance error term (see, e.g., Raykov et al., 2016). We also note that for $q = 3$ criterion indicators model (7) is equivalent to a corresponding first-order factor model with correlated traits. For the last model, the criterion estimation procedure in Raykov and Marcoulides (2011, chap. 8) is straight-forwardly utilizable with respect to any of these traits in case of a latent criterion with multiple indicators; if only a single criterion measure is available then that is error-free, the same procedure can also be used after the minor modification mentioned above in this footnote.

References

Adorno, T. W., Frenkel-Brunswik, E., Levinson, D., & Sanford, N. (1950). *The authoritarian personality*. New York, NY: Harper.

Altemeyer, B. (1981). *Right-wing authoritarianism*. Winnipeg, Manitoba, Canada: University of Manitoba Press.

Banaszak, L. A., & Plutzer, E. (1993). The social bases of feminism in the European community. *Public Opinion Quarterly*, *57*, 29-53.

Beierlein, C., Asbrock, F., Kauff, M., & Schmidt, P. (2014). Die Kurzskala Autoritarismus (KSA-3): Ein ökonomisches Messinstrument zur Erfassung dreier Subdimensionen autoritärer Einstellungen [A short scale of authoritarianism (KSA-3): An inventory for the measurement of the three dimensions of authoritarian attitudes]. In D. Danner & A. Glöckner-Rist (Eds.), *Zusammenstellung sozialwissenschaftlicher Items und Skalen* [*Collection of Items and Scales for the Social Sciences*]. Mannheim: GESIS. doi:10.6102/zis228. Retrieved from http://zis.gesis.org/skala/Beierlein-Asbrock-Kauff-Schmidt-Kurzskala-Autoritarismus-%28KSA-3%29.

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

Braun, M. (2014). Gender-role attitudes (ISSP 94). In D. Danner & A. Glöckner-Rist (Eds.), *Zusammenstellung sozialwissenschaftlicher Items und Skalen* [*Collection of Items and Scales for the Social Sciences*]. Mannheim: GESIS. doi:10.6102/zis223. Retrieved from http://zis.gesis.org/skala/Braun-Gender-role-attitudes-%28ISSP-94%29.

Crawford, J. T., & Pilanski, J. M. (2014). Political intolerance, right and left. *Political Psychology*, *35*, 841-851.

Crocker, L., & Algina, J. (2006). *Introduction to classical and modern test theory*. Fort Worth, TX: Harcourt College.

DiStefano, C. (2002). The impact of categorization with confirmatory factor analysis. *Structural Equation Modeling*, *9*, 327-346.

Duckitt, J., & Bizumic, B. (2013). Multidimensionality of right-wing authoritarian attitudes: authoritarianism-conservatism-traditionalism. *Political Psychology*, *34*, 841-862.

Enders, C. K. (2010). *Applied missing data analysis*. New York, NY: Guilford Press.

Ferrando, P. J., & Lorenzo-Seva, U. (2010). Acquiescence as a source of bias and model and person fit: A theoretical and empirical analysis. *British Journal of Mathematical and Statistical Psychology*, *63*, 427-428.

GESIS–Leibniz-Institute for the Social Sciences. (2016). German General Social Survey (ALLBUS)–Cumulation 1980-2014. GESIS Datenarchiv, Köln. ZA4584 Datenfile Version 1.0.0, doi:10.4232/1.12574

Harman, H. H. (1976). *Modern factor analysis*. Chicago, IL: University of Chicago Press.

Haidt, J., Graham, J., & Joseph, C. (2009). Above and below left–right: Ideological narratives and moral foundations. *Psychological Inquiry*, *20*, 110-119.

Knight, K. (1993). Liberalism and conservatism. In J. Robinson, P. Shaver, & L. Wrightsman (Eds.), *Political attitudes* (pp. 59-158). San Diego, CA: Academic Press.

Leone, L., Desimoni, M., & Chirumbolo, A. (2014). Interest and expertise moderate the relationship between right-wing attitudes, ideological self-placement and voting. *European Journal of Personality*, *28*, 2-13.

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

Messick, S. (1995). Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, *50*, 741-749.

Muthén, B. O. (2002). Beyond SEM: General latent variable modeling. *Behaviormetrika*, *29*, 81-117.

Muthén, L. K., & Muthén, B. (2017). *Mplus user's guide*. Los Angeles, CA: Muthén & Muthén.

Raykov, T. (2012). Scale development using structural equation modeling. In R. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 472-492). New York, NY: Guilford Press.

Raykov, T., & Marcoulides, G. A. (2006). *A first course in structural equation modeling*. Mahwah, NJ: Lawrence Erlbaum.

Raykov, T., & Marcoulides, G. A. (2011). *Introduction to psychometric theory*. New York, NY: Taylor & Francis.

Raykov, T., Marcoulides, G. A., Gabler, S., & Lee, Y. (2017). Testing criterion correlations with scale component measurement errors using latent variable modeling. *Structural Equation Modeling*, *24*, 468-474. doi:10.1080/10705511.2015.1136220

Raykov, T., Marcoulides, G. A., & Patelis, T. (2015). The importance of the assumption of uncorrelated errors in psychometric theory. *Educational and Psychological Measurement*, *75*, 634-647.