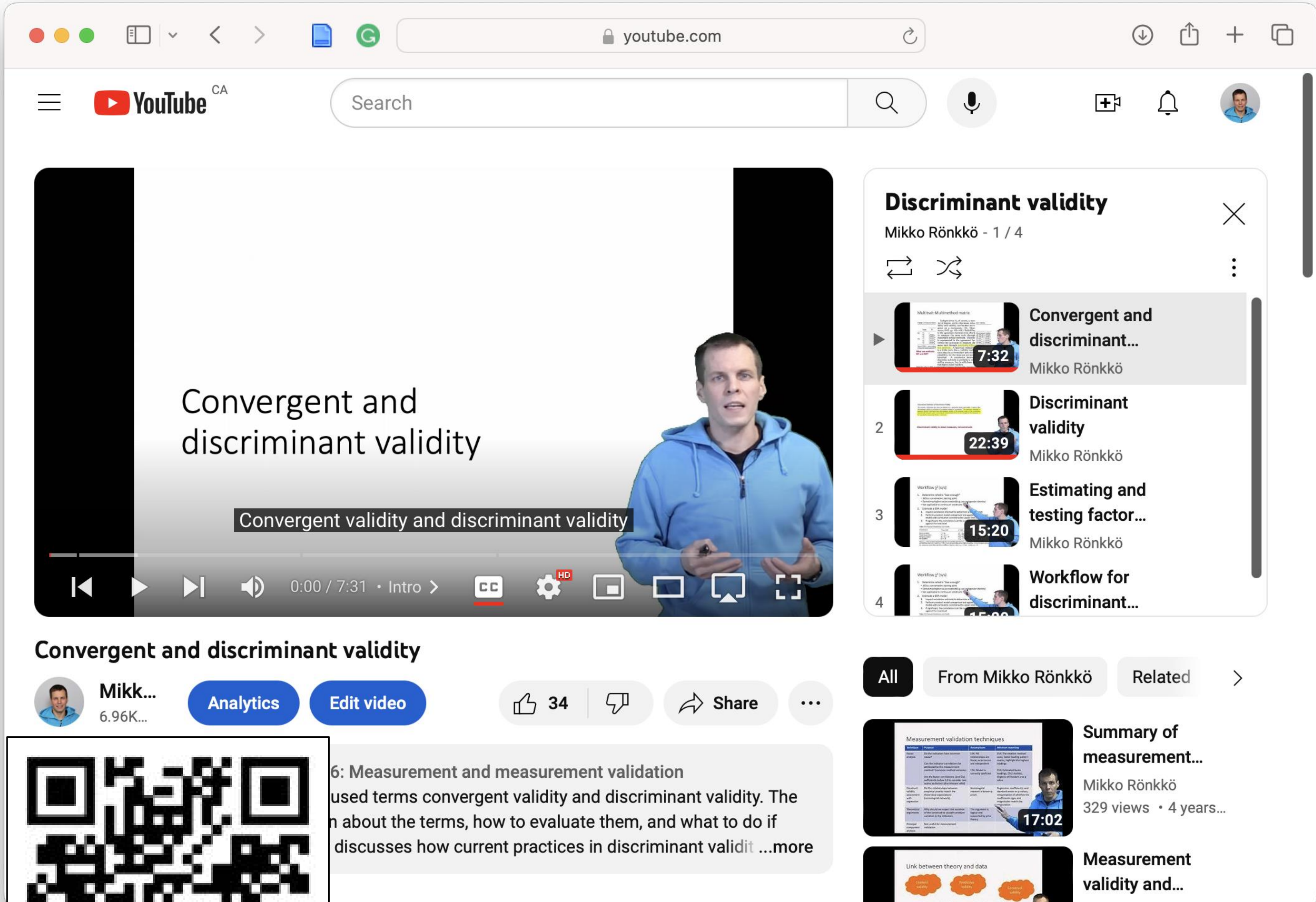
Convergent and discriminant validity

Mikko Rönkkö

Jyväskylä University School of Business and Economics, Finland


youtube.com
YouTube CA
Search
Convergent and discriminant validity
Convergent validity and discriminant validity
0:00 / 7:31 • Intro
Discriminant validity
Mikko Rönkkö - 1 / 4
Convergent and discriminant...
Mikko Rönkkö
7:32
Discriminant validity
Mikko Rönkkö
22:39
Estimating and testing factor...
Mikko Rönkkö
15:20
Workflow for discriminant...
Convergent and discriminant validity
Mikk...
6.96K...
Analytics
Edit video
34
Share
All
From Mikko Rönkkö
Related
6: Measurement and measurement validation
used terms convergent validity and discriminant validity. The
n about the terms, how to evaluate them, and what to do if
discusses how current practices in discriminant validit ...more
Summary of measurement...
Mikko Rönkkö
329 views • 4 years...
17:02
Measurement validity and...

Organizational Research Methods
1-42

Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/1094428120968614
journals.sagepub.com/home/orm
SAGE

An Updated Guideline for Assessing Discriminant Validity

Mikko Rönkkö[1] **and Eunseong Cho**[2]

Abstract

Discriminant validity was originally presented as a set of empirical criteria that can be assessed from multitrait-multimethod (MTMM) matrices. Because datasets used by applied researchers rarely lend themselves to MTMM analysis, the need to assess discriminant validity in empirical research has led to the introduction of numerous techniques, some of which have been introduced in an ad hoc manner and without rigorous methodological support. We review various definitions of and techniques for assessing discriminant validity and provide a generalized definition of discriminant validity based on the correlation between two measures after measurement error has been considered. We then review techniques that have been proposed for discriminant validity assessment, demonstrating some problems and equivalencies of these techniques that have gone unnoticed by prior research. After conducting Monte Carlo simulations that compare the techniques, we present techniques called CI_{CFA}(sys) and w^2(sys) that applied researchers can use to assess discriminant validity.

Keywords

discriminant validity, Monte Carlo simulation, measurement, confirmatory factor analysis, validation, average variance extracted, heterotrait-monotrait ratio, cross-loadings

Organizational Research Methods
15(1) 17-32

DOI: 10.1177/1094428110392383
http://orm.sagepub.com

Understanding the Impact of Convergent Validity on Research Results

Kevin D. Carlson[1] and Andrew O. Herdman[2]

Abstract

Using different measures of constructs in research to develop robust evidence of relationships and effects is seen as good methodological practice. This assumes these measures possess high convergent validity. However, proxies—alternative measures of the same construct—are rarely perfectly convergent. Although some convergence is preferred to none, this study demonstrates that even modest departures from perfect convergent validity can result in substantial differences in the magnitudes of findings, creating challenges for the accumulation and interpretation of research. Using data from published research, the authors find that substantial differences in findings between studies using desired and proxy variables occur even at levels of convergent validity as high as $r = .85$. Implications of using measures with less-than-ideal convergent validity for the interpretability of research results are examined. Convergent validities above $r = .70$ are recommended, whereas those below $r = .50$ should be avoided. Researchers are encouraged to develop and report convergent validity data.

Keywords

field research methods, research design, measurement design, content validity, reliability and

Agenda

1. Introduction to convergent and discriminant validity
2. How does convergent validity relate to reliability
3. When do you need to assess discriminant and convergent validity
4. Statistical techniques for assessing convergent and discriminant validity
5. Workflow and reporting

Introduction to convergent and discriminant validity

Convergent and discriminant validity

> The measurement model can be used to evaluate discriminant validity. Constructs demonstrate discriminant validity if the variance extracted for each is higher than the squared correlation between the constructs (Fornell and Larcker 1981). We examined each pair of constructs in our measurement model and found that all demonstrate discriminant validity. Convergent validity is also evident: positive correlations exist among the three social capital constructs, as is expected for constructs representing different dimensions of the same underlying concept. Table 2 reports means, standard deviations, ranges, and correlations for the variables of the study.

- Introduced by Campbell and Fiske (1959)
  - No definition
  - Empirical tests using MTMM correlations
  - Convergent validity refers to whether indicators that are supposed to measure the same thing correlate
  - Discriminant validity refers to whether indicators that measure different things do not correlate too strongly

Yli-Renko, H., Autio, E., & Sapienza, H. J. (2001). Social capital, knowledge acquisition, and knowledge exploitation in young technology-based firms. *Strategic Management Journal*, *22*(6–7), 587–613. https://doi.org/10.1002/smj.183 p. 602
Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*(2), 81–105.

Multitrait-Multimethod matrix

Table 3. Multitrait-Multimethod Correlation Matrix and Original Criteria for Discriminant Validity.

		Method M1			Method M2			
	Traits	T1	T2	T3	T1	T2	T3	
M1	T1	1						Discriminant Validity:
	T2	$HTMM_{11}$	1					All MTHM > HTHM
	T3	$HTMM_{12}$	$HTMM_{13}$	1				All MTHM > HTMM
M2	T1	$MTHM_{21}$	$HTHM_{24}$	$HTHM_{27}$	1			$HTMM_{11} \approx HTMM_{31}$
	T2	$HTHM_{22}$	$MTHM_{25}$	$HTHM_{28}$	$HTMM_{31}$	1		$HTMM_{12} \approx HTMM_{32}$
	T3	$HTHM_{23}$	$HTHM_{26}$	$MTHM_{29}$	$HTMM_{32}$	$HTMM_{33}$	1	$HTMM_{13} \approx HTMM_{33}$

Note: HTMM = same method and different traits (heterotrait-monomethod); MTHM = different methods and same trait (monotrait-heteromethod); HTHM = different methods and different traits (heterotrait-heteromethod).

Rönkkö, M., & Cho, E. (2022). An updated guideline for assessing discriminant validity. *Organizational Research Methods*, *25*(1), 6–14. https://doi.org/10.1177/1094428120968614

Modern discriminant validity

Table 2. Definitions of Discriminant Validity in Existing Studies.

Category	Definition/Description of Technique
1: True or estimated correlation between constructs[a]	"[T]he degree to which the absolute value of the correlation between the two constructs differ from one." (Reichardt & Coleman, 1995, p. 516)
	"Evidence of discriminant validity exists if other constructs do not correlate strongly enough with the construct of interest to suggest that they measure the same construct." (McKenny et al., 2013, p. 156)
	"Discriminant validation implies that correlation between traits is low. If both traits were identical, the correlation between the trait factors would be near one." (Kenny, 1976, p. 251)
	"[D]iscriminant validity exists when estimates of the trait correlations were two or more standard errors below 1.0." (Schmitt & Stults, 1986, p. 18)
	"[D]iscriminant validity consists of demonstrating that the true correlation of [two traits] is meaningfully less than unity." (Werts & Linn, 1970, p. 208)
2: Correlation between measures	"[A] test [should] not correlate too highly with measures from which it is supposed to differ." (Campbell, 1960, p. 548)
	"[A test] correlates less well or not all with tests with which theory implies it should not correlate well." (McDonald, 1985, p. 220)
	"[T]he extent to which measures of theoretically distinct constructs are unrelated empirically to one another." (J. A. Shaffer et al., 2016, p. 82)
	"[I]f two or more concepts are unique, then valid measures of each should not correlate too highly." (Bagozzi et al., 1991, p. 425)
	"[T]he degree of divergence among indicators that are designed to measure different constructs." (Hamann et al., 2013, p. 72)
	"[T]he degree to which measures of distinct concepts differ." (Bagozzi & Phillips, 1982, p. 469)
	"Measures of different attributes should . . . not correlated to an extremely high degree." (Nunnally & Bernstein, 1994, p. 93)
	"[A] measure of a construct is unrelated to indicators of theoretically irrelevant constructs in the same domain." (Strauss & Smith, 2009, p. 1)
3: Correlation between measure and other construct	"[D]iscriminant validity is shown when each measurement item correlates weakly with all other constructs except for the one to which it is theoretically associated." (Gefen & Straub, 2005, p. 92)
	"Discriminant validity is inferred when scores from measures of different constructs do not converge. It thus provides information about whether scores from a measure of a construct are unique rather than contaminated by other constructs." (Schwab, 2013, p. 33)
4: Combination of categories 1 and 3.	"[T]he item's . . . loading on constructs other than the intended one is relevant to discriminant validity . . . . At the level of the constructs, this correlation tells us about discriminant validity." (John & Benet-Martínez, 2000, p. 359)
	Voorhees et al. (2016) classified discriminant validity into the construct-level (i.e., low correlation) and the item-level (i.e., absence of cross-loading).

a. Many articles in this category are ambiguous on whether discriminant validity is a property of a construct or a property of a scale from which construct correlation is estimated.

Generalized Definition of Discriminant Validity

We present a definition that does not depend on a particular model and makes it explicit that discriminant validity is a feature of a measure instead of a construct:[2] *Two measures intended to measure distinct constructs have discriminant validity if the absolute value of the correlation between the measures after correcting for measurement error is low enough for the measures to be regarded as measuring distinct constructs.*

Discriminant validity is about measures, not constructs

Factor model is not a part of the definition

Rönkkö, M., & Cho, E. (2022). An updated guideline for assessing discriminant validity. *Organizational Research Methods*, *25*(1), 6–14. https://doi.org/10.1177/1094428120968614

Generalized Definition of Discriminant Validity

We present a definition that does not depend on a particular model and makes it explicit that discriminant validity is a feature of a measure instead of a construct:[2] *Two measures intended to measure distinct constructs have discriminant validity if the absolute value of the correlation between the measures after correcting for measurement error is low enough for the measures to be regarded as measuring distinct constructs.*

This definition encompasses the early idea that even moderately high correlations between distinct measures can invalidate those measures if measurement error is present (Thorndike, 1920), which serves as the basis of discriminant validity (Campbell & Fiske, 1959). The definition can also be applied on both the scale level and the scale-item level. Consider the proposed definition in the context of the common factor model:

$$\Sigma = \Lambda\Phi\Lambda' + \Theta \tag{1}$$

where Σ is the interitem correlation matrix; Φ is the factor correlation matrix, where all correlations are assumed to be positive for simplicity; Λ is a factor pattern (loading) matrix; and Θ is the item error covariance matrix. Within this context, our definition can be understood in two equivalent ways:

$$\Sigma_{i,j} \ll \Lambda J \Lambda'_{i,j} \tag{2}$$

$$(\Lambda'\Lambda)^{-1}\Lambda'(\Sigma - \Theta)\Lambda(\Lambda'\Lambda)^{-1}_{k,l} \ll 1 \tag{3}$$

where J is a unit matrix (a matrix of ones) and $\ll$ denotes much less than. Equation 2 is an item-level comparison (category 2 in Table 2), where the correlation between items i and j, which are designed to measure different constructs, is compared against the implied correlation when the items depend on perfectly correlated factors but are not perfectly correlated because of measurement error. Equation 3 shows an equivalent scale-level comparison (part of category 1 in Table 2) focusing on two distinct scales k and l. The factor correlations are solved from the interitem correlations by multiplying with left and right inverses of the factor pattern matrix to correct for measurement error and are then compared against a perfect correlation. Generalizing beyond the linear common factor model, Equation 3 can be understood to mean that *two scales intended to measure distinct constructs have discriminant validity if the absolute value of the correlation between two latent variables estimated from the scales is low enough for the latent variables to be regarded as representing distinct constructs*.

Rönkkö, M., & Cho, E. (2022). An updated guideline for assessing discriminant validity. *Organizational Research Methods*, *25*(1), 6–14. https://doi.org/10.1177/1094428120968614

Generalized Definition of Discriminant Validity

We present a definition that does not depend on a particular model and makes it explicit that discriminant validity is a feature of a measure instead of a construct:[2] *Two measures intended to measure distinct constructs have discriminant validity if the absolute value of the correlation between the measures after correcting for measurement error is low enough for the measures to be regarded as measuring distinct constructs.*

What is is "low enough"?

Magnitude of the Discriminant Validity Correlations

In the discriminant validity literature, high correlations between scales or scale items are considered problematic. However, the literature generally has not addressed what is high enough beyond giving rule of thumb cutoffs (e.g., 85). Our definition of discriminant validity suggests that the magnitude of the estimated correlation depends on the correlation between the constructs, the measurement process, and the particular sample, each of which has different implications on what level should be considered high. To warn against mechanical use, we present a scenario where high correlation does not invalidate measurement and a scenario where low correlation between measures does not mean that they measure distinct constructs.

A large correlation does not always mean a discriminant validity problem if one is expected based on theory or prior empirical observations. For example, the correlation between biological sex and gender identity can exceed .99 in the population.[17] However, both variables are clearly distinct: sex is a biological property with clear observable markers, whereas gender identity is a psychological construct. These two variables also have different causes and consequences (American Psychological Association, 2015), so studies that attempt to measure both can lead to useful policy implications. In cases such as this where the constructs are well defined, large correlations should be tolerated when expected based on theory and prior empirical results. Of course, large samples and precise measurement would be required to ensure that the constructs can be distinguished empirically (i.e., are empirically distinct).

Rönkkö, M., & Cho, E. (2022). An updated guideline for assessing discriminant validity. *Organizational Research Methods*, *25*(1), 6–14. https://doi.org/10.1177/1094428120968614

Generalized Definition of Discriminant Validity

We present a definition that does not depend on a particular model and makes it explicit that discriminant validity is a feature of a measure instead of a construct:[2] *Two measures intended to measure distinct constructs have discriminant validity if the absolute value of the correlation between the measures after correcting for measurement error is low enough for the measures to be regarded as measuring distinct constructs.*

What is is "low enough"?

A small or moderate correlation (after correcting for measurement error) does not always mean that two measures measure concepts that are distinct. For example, consider two thermometers that measure the same temperature, yet one is limited to measuring only temperatures above freezing, whereas the other can measure only temperatures below freezing. While both measure the same quantity, they are correlated only by approximately .45 because the temperature would always be out of the range of one of the thermometers that would consequently display zero centigrade.[18] In the social sciences, a well-known example is the measurement of happiness and sadness, two constructs that can be thought of as opposite poles of mood (D. P. Green et al., 1993; Tay & Jebb, 2018). Consequently, any evaluation of the discriminant validity of scales measuring two related constructs must precede the theoretical consideration of the existence of a common continuum. If this is the case, the typical discriminant validity assessment techniques that are the focus of our article are not directly applicable, but other techniques are needed (Tay & Jebb, 2018).

As the two examples show, a moderately small correlation between measures does not always imply that two constructs are distinct, and a high correlation does not imply that they are not. Like any validity assessment, discriminant validity assessment requires consideration of context, possibly relevant theory, and empirical results and cannot be reduced to a simple statistical test and a cutoff no matter how sophisticated. These considerations highlight the usefulness of the continuous interpretation of discriminant validity evidence.

Rönkkö, M., & Cho, E. (2022). An updated guideline for assessing discriminant validity. *Organizational Research Methods*, *25*(1), 6–14. https://doi.org/10.1177/1094428120968614

Summary of discriminant validity

Two measures intended to measure distinct constructs have discriminant validity if the absolute value of the correlation between the measures after correcting for measurement error is low enough for the measures to be regarded as measuring distinct constructs.

Table 12. Proposed Classification and Cutoffs.

Classification	CI_{CFA} (sys)	χ^2 (sys)
Severe problem	$1 \leq UL$	$\chi^2_1 - \chi^2_{org} < 3.84$
Moderate problem	$.9 \leq UL < 1$	Not "Marginal problem" AND $\chi^2_1 - \chi^2_{org} > 3.84$
Marginal problem	$.8 \leq UL < .9$	Not "No problem" AND $\chi^2_{.9} - \chi^2_{org} > 3.84$
No problem	$UL < .8$	$\rho_{CFA} < .8$ AND $\chi^2_{.8} - \chi^2_{org} > 3.84$

Note: ρ_{CFA} is the correlation obtained using CFA, UL is the 95% upper limit of ρ_{CFA} when $\rho_{CFA} > 0$, and the absolute value of the 95% lower limit of ρ_{CFA} when $\rho_{CFA} < 0$, χ^2_{org} is the chi-square value of the original model, and χ^2_c is the chi-square value of the comparison model where the focal correlation is fixed to c when $\rho_{CFA} > 0$ and $-c$ when $\rho_{CFA} < 0$.

Any cutoff is ultimately arbitrary

Modern convergent validity and reliability

Multitrait-Multimethod matrix

Table 3. Multitrait-Multim … inant Validity.

	Traits	T1
M1	T1	1
	T2	$HTMM_{11}$
	T3	$HTMM_{12}$
M2	T1	$MTHM_{21}$
	T2	$HTHM_{22}$
	T3	$HTHM_{23}$

Discriminant Validity:
All MTHM > HTHM
All MTHM > HTMM
$HTMM_{11} \approx HTMM_{31}$
$HTMM_{12} \approx HTMM_{32}$
$HTMM_{13} \approx HTMM_{33}$

Note: HTMM = same methoc … it methods and same trait (monotrait-heteromethod); H … ıethod).

Independence is, of course, a matter of degree, and in this sense, reliability and validity can be seen as regions on a continuum. (Cf. Thurstone, 1937, pp. 102–103.) Reliability is the agreement between two efforts to measure the same trait through maximally similar methods. Validity is represented in the agreement between two attempts to measure the same trait through maximally different methods. A split-half reliability is a little more like a validity coefficient than is an immediate test-retest reliability, for the items are not quite identical. A correlation between dissimilar subtests is probably a reliability measure, but is still closer to the region called validity.

What are methods M1 and M2?

Rönkkö, M., & Cho, E. (2020). An updated guideline for assessing discriminant validity. *Organizational Research Methods*. https://doi.org/10.1177/1094428120968614
Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*(2), 81–105. p. 83

Example convergent validity and reliability

Reliability

Convergent validity

First measure

Second measure

Please estimate the weight of the person you see

Same correlation can be either convergent validity or reliability evidence, depending on measures

Assumption: lack of reliability the only reason measures not perfectly correlated

How strongly should distinct measures correlate?

Carlson and Herdman 27

Table 2. Analysis of Absolute Differences in r_{ay} and r_{by} Across Correlation Pairs at Different Levels of Convergent Validity

Recom-
mended
range

Convergent Validity	*N*	*M*	*SD*	Max	Distribution of Absolute Differences					
					.00-.02	.03-.05	.06-.10	.11-.20	.21-.30	>.30
.95	237	.02	.03	.26	174 (73)	43 (18)	15 (06)	4 (02)	1 (00)	0 (00)
.90	312	.04	.04	.42	122 (39)	115 (37)	63 (20)	11 (04)	0 (00)	1 (00)
.85	325	.06	.11	1.21	113 (35)	79 (24)	87 (27)	40 (12)	2 (01)	4 (01)
.80	325	.09	.08	.64	85 (25)	72 (22)	75 (23)	72 (22)	10 (03)	11 (03)
.75	431	.08	.08	.88	96 (22)	98 (23)	116 (27)	94 (22)	21 (05)	4 (01)
.70	345	.09	.09	.76	84 (24)	67 (19)	92 (71)	71 (21)	19 (06)	12 (03)
.60	384	.11	.14	1.49	67 (18)	86 (24)	76 (21)	83 (23)	37 (10)	15 (04)
.50	377	.13	.12	.95	53 (14)	67 (17)	98 (25)	103 (26)	40 (10)	27 (07)
.30	347	.16	.16	.86	40 (11)	44 (12)	78 (22)	96 (27)	49 (14)	51 (14)
.10	525	.18	.18	1.18	51 (10)	70 (14)	78 (16)	111 (23)	82 (17)	100 (20)

Note: *N* is the number of correlation pairs examined; *M* is the mean and *SD* is the standard deviation of differences in the magnitude of r_{ay} and r_{by}. *Max* is the largest value found between the two correlations. In the columns labeled *distribution of absolute values,* the number is the actual number of correlation pairs where the absolute value of the difference falls in the identified range. The number in parentheses is the percentage of the total *N* represented by the preceding number.

Carlson, K. D., & Herdman, A. O. (2012). Understanding the Impact of Convergent Validity on Research Results. *Organizational Research Methods*, *15*(1), 17–32. https://doi.org/10.1177/1094428110392383

When to assess convergent and discriminant validity

Commonly requested by reviewers

Convergent and discriminant validity. Another commonly critiqued aspect of measurement was a lack of convergent and discriminant evidence provided by the authors. This issue was raised in letters corresponding to a little over half of manuscripts ($n = 39$; 56.5%). Editors and reviewers questioned the convergent and discriminant validity of constructs when authors did not provide evidence of validity. For example, one reviewer's comment exemplifying this concern read, "Given the high correlations among the three forms of conflict, I would like to see evidence of discriminant validity (i.e., CFA)."

When measures were related, reviewers and editors often suggested that they might be indicators of the same construct. One reviewer reflected these thoughts for a paper on climate, "The two ... scales share a manifest level correlation of .66 .... Maybe the two scales should be used as two indicators of a common construct." Some of reviewers' and editors' requests for factor analyses and structural equation modeling (as discussed further in the Data Analytic Errors section) were, in part, an attempt to make sure that the constructs were distinct, such as one reviewer's request that the authors "at least employ [an] exploratory factor analysis with all of the studies' items to provide additional evidence that the measures are distinct." Additionally, reasons provided by two reviewers to conduct a CFA were to "make sure the items load on the measures appropriately" and to "confirm that [constructs] are in fact different and appropriately measured."

But most studies do not have multiple methods that are maximally different?

Green, J. P., Tonidandel, S., & Cortina, J. M. (2016). Getting through the gate: Statistical and methodological issues raised in the reviewing process. *Organizational Research Methods*, *19*(3), 402–432. https://doi.org/10.1177/1094428116631417 p. 412

When to report discriminant and convergent validity evidence

Convergent validity

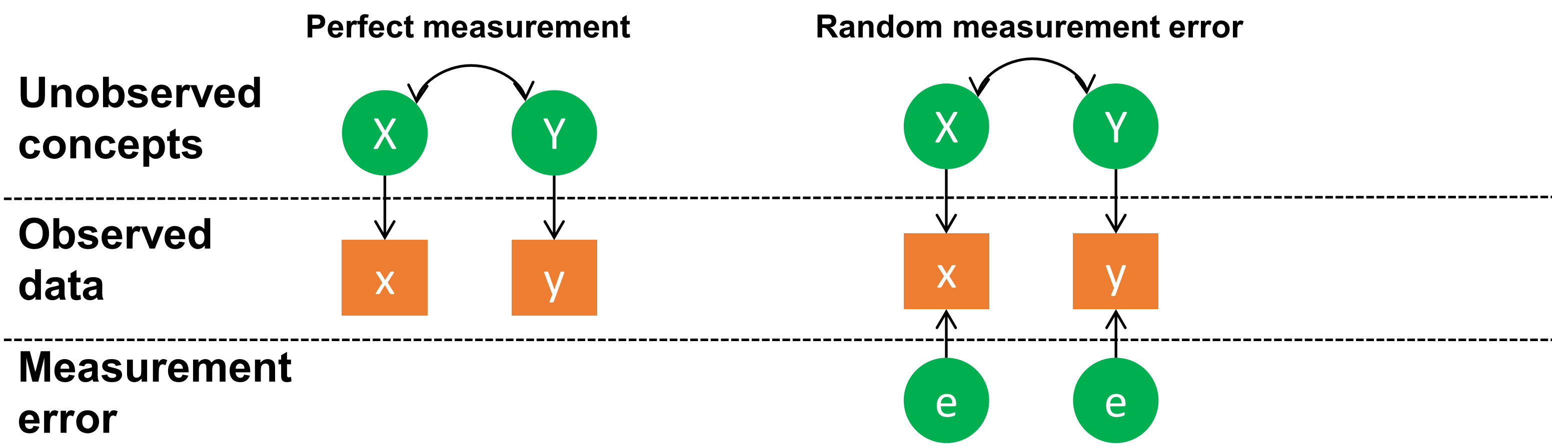
- When distinct measures are used
  - e.g., Validating a perceptual measure of firm performance with an accounting measure
  - When proxy measures are used
- Not when similar measures are used
  - e.g., A multiple-item scale
- Not if a correlation is used in a unidimensional reliability measure (e.g., alpha)

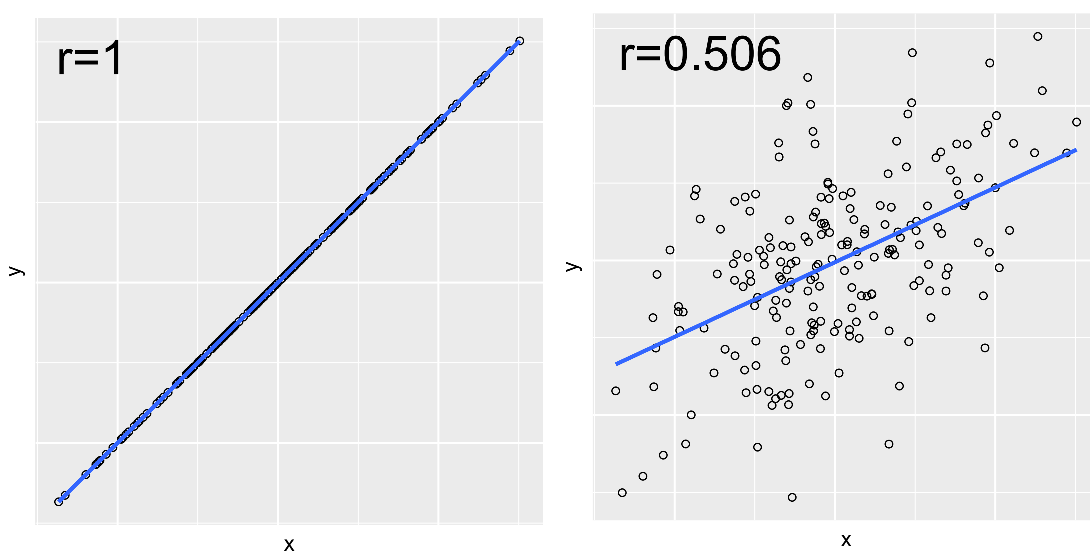
Discriminant validity

- When there is a concern that measures intended to capture different things actually capture the same thing
- When correlations between scales are high (e.g. factor correlations >.50)

Overview of discriminant validity assesment techniques

Effect of random measurement error





Variances of error term(s) e and latent variables X and Y are equal

Table 1. Techniques Used to Assess Discriminant Validity in *AMJ*, *JAP*, and *ORM*.

	AMJ ($n = 27$)		*JAP* ($n = 73$)		*ORM* ($n = 5$)		
Techniques using correlation estimates							
Scale score correlation (ρ_{SS})	7	25.9%	8	11.0%	3	60.0%	
Factor correlation (ρ_{CFA})	0	0.0%	2	2.7%	1	20.0%	CI_{CFA}
Disattenuated correlation (ρ_{DTR})	0	0.0%	1	1.4%	1	20.0%	
Techniques to compare AVE to a certain value							
AVE_{CFA} vs. Square of ρ_{CFA} (AVE/SV_{CFA})	2	7.4%	4	5.5%	1	20.0%	
AVE_{CFA} vs. Square of ρ_{SS} (AVE/SV_{SS})	1	3.7%	1	1.4%	0	0.0%	
AVE_{CFA} vs. .5	2	7.4%	1	1.4%	0	0.0%	
AVE_{PLS} vs. Square of ρ_{SS}	2	7.4%	0	0.0%	0	0.0%	
Techniques to show low cross-loadings							
CFA (structure coefficients)	2	7.4%	0	0.0%	0	0.0%	
Exploratory factor analysis	1	3.7%	0	0.0%	0	0.0%	
Techniques using fit indices of CFA models							
No comparison (only the proposed model)	3	11.1%	1	1.4%	0	0.0%	
Compared with nested models with fewer factors (χ^2(merge))	8	29.6%	43	58.9%	1	20.0%	
Compared with model with fixed correlation of 1 ($\chi^2(1)$)	4	14.8%	1	1.4%	2	40.0%	$\chi^2(1)$
Compared with model with fixed correlation of 1 (CFI(1))	0	0.0%	0	0.0%	1	20.0%	
Techniques requiring multiple measurement methods							
GCES approach	0	0.0%	0	0.0%	1	20.0%	χ^2(ref)
MTMM approach	0	0.0%	1	1.4%	0	0.0%	
Generalizability theory approach	0	0.0%	1	1.4%	0	0.0%	
Techniques that are difficult to classify							
CFA results not presented in detail	1	3.7%	4	5.5%	0	0.0%	
No clear evidence provided	0	0.0%	3	4.1%	0	0.0%	
Comparison with existing research results	0	0.0%	2	2.7%	0	0.0%	
Experimental results as expected	0	0.0%	2	2.7%	0	0.0%	

Note: The sum exceeds 100% because some studies use multiple techniques. *AMJ* = *Academy of Management Journal*; *JAP* = *Journal of Applied Psychology*; *ORM* = *Organizational Research Methods*; CFA = confirmatory factor analysis; AVE = average variance extracted; ρ_{CFA} = factor correlation obtained from CFA; ρ_{TR} = disattenuated correlation using tau-equivalent reliability; AVE_{CFA} = AVE obtained from CFA; AVE_{PLS} = AVE obtained from partial least squares; GCES = generalized coefficient of equivalence and stability; MTMM = multitrait-multimethod. For a detailed description of the symbols, see Table 4.

Rönkkö, M., & Cho, E. (2022). An updated guideline for assessing discriminant validity. *Organizational Research Methods*, *25*(1), 6–14. https://doi.org/10.1177/1094428120968614

Estimating factor correlation with CFA

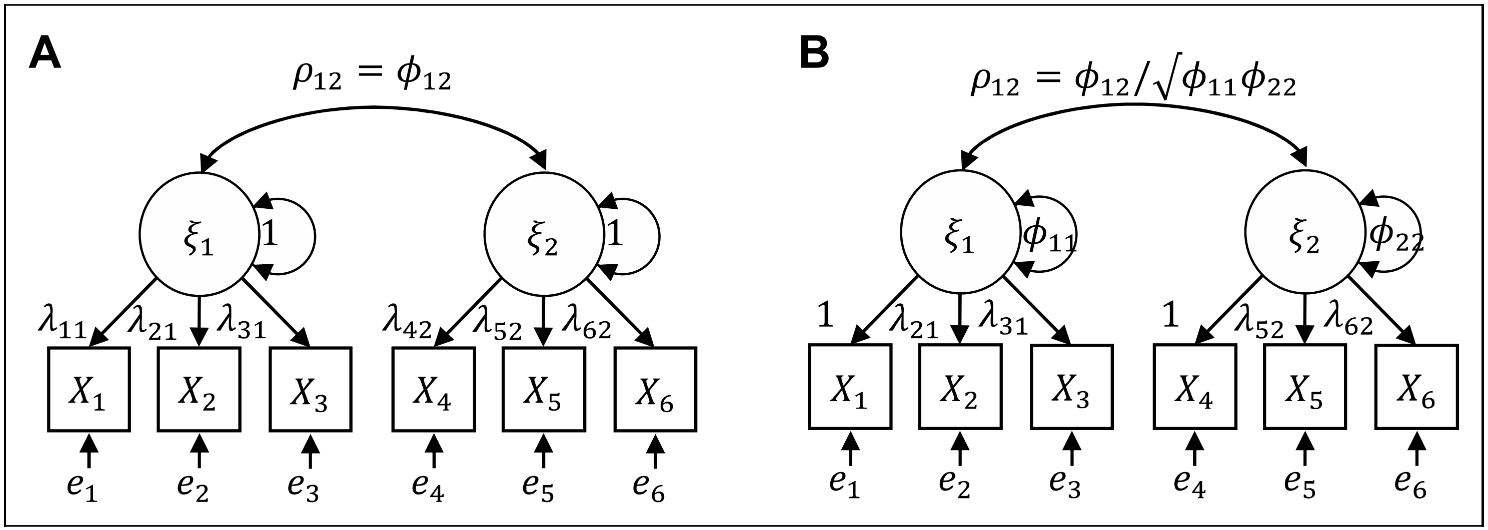


Figure 2. Factor correlation estimation. (A) Fixing the variances of factors to unity (i.e., not using the default option). (B) Fixing one of the loadings to unity (i.e., using the default option).

CI_{CFA} Standardization after estimation works

$\chi^2(1)$ and $\chi^2(\text{ref})$ estimated model must be standardized

Rönkkö, M., & Cho, E. (2022). An updated guideline for assessing discriminant validity. *Organizational Research Methods*, *25*(1), 6–14. https://doi.org/10.1177/1094428120968614

Nested model comparison

1. Estimate a factor model where all factor correlations are freely estimated (unconstrained model)
2. Estimate constrained model A ($\chi^2(1)$ and χ^2(ref)) or C (χ^2(merge))
3. Compare constrained model against unconstrained model
4. Repeat for each factor pair

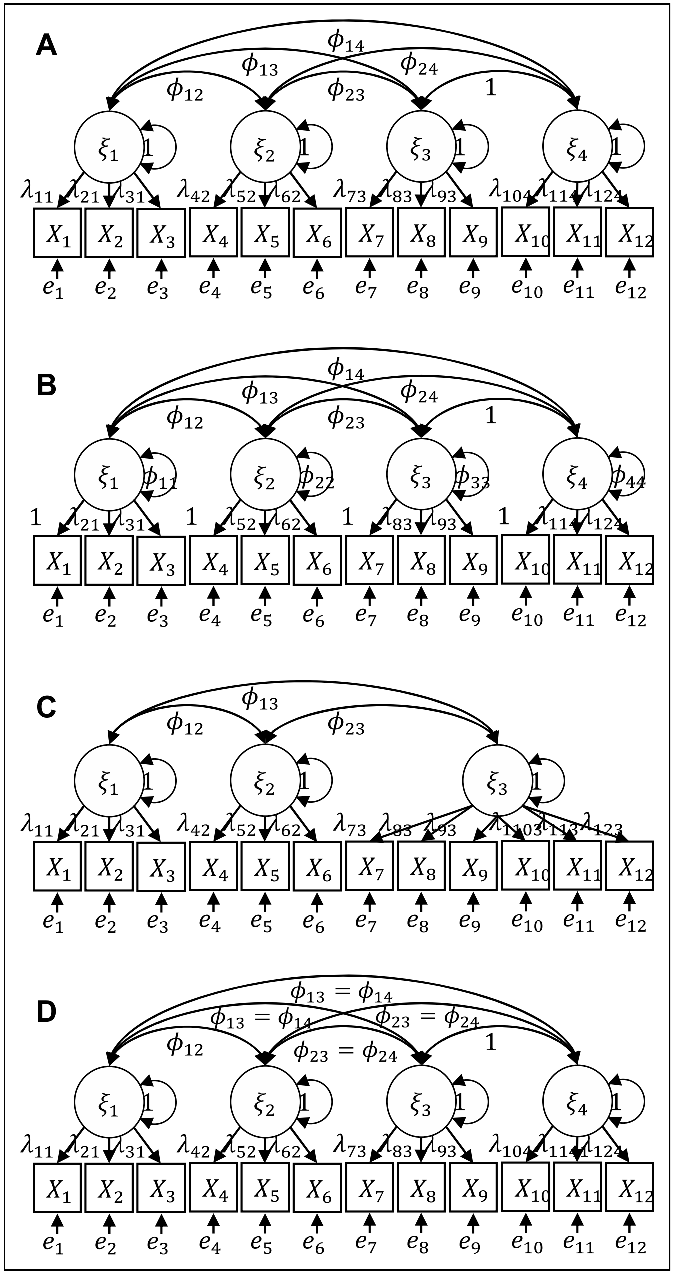


Figure 5. (A) constrained model for $\chi^2(1)$, (B) common misuse of $\chi^2(1)$, (C) constrained model for χ^2(merge), (D) model equivalent to C.

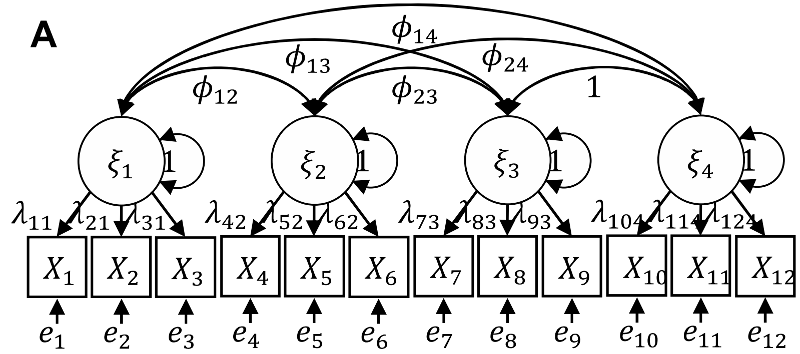
Rönkkö, M., & Cho, E. (2022). An updated guideline for assessing discriminant validity. *Organizational Research Methods*, *25*(1), 6–14. https://doi.org/10.1177/1094428120968614

```
> library(lavaan); library(semTools)

> fit <- cfa('visual  =~ x1 + x2 + x3
              textual =~ x4 + x5 + x6
              speed   =~ x7 + x8 + x9 ',
              data = HolzingerSwineford1939)

> discriminantValidity(fit)

Some of the latent variable variances are estimated instead of fixed to 1. The
model is re-estimated by scaling the latent variables by fixing their variances and
freeing all factor loadings.
      lhs op     rhs  est ci.lower ci.upper Df  Chisq Chisq diff Df diff Pr(>Chisq)
1  visual ~~ textual 0.46     0.33     0.58 25 152.23      66.92       1          0
2  visual ~~   speed 0.47     0.33     0.61 25 124.02      38.71       1          0
3 textual ~~   speed 0.28     0.15     0.42 25 200.42     115.12       1          0

> discriminantValidity(fit, merge = TRUE)

Some of the latent variable variances are estimated instead of fixed to 1. The
model is re-estimated by scaling the latent variables by fixing their variances and
freeing all factor loadings.
      lhs op     rhs  est ci.lower ci.upper Df  Chisq Chisq diff Df diff Pr(>Chisq)
1  visual ~~ textual 0.46     0.33     0.58 26 181.34      96.03       2          0
2  visual ~~   speed 0.47     0.33     0.61 26 151.47      66.16       2          0
3 textual ~~   speed 0.28     0.15     0.42 26 236.09     150.79       2          0
```

Summary of discriminant validity techniques

1. Estimate a factor model where all factors correlations are freely estimated, scale by fixing variances
2. Interpret confidence intervals CI_{CFA} OR estimate a series of constrained models and use nested model test: $\chi^2(1)$ or $\chi^2(\text{ref})$

A



Online supplements of Rönkkö and Cho (2020) provide tutorials for Stata, R (Lavaan), Amos, LISREL, and Mplus.

Rönkkö, M., & Cho, E. (2022). An updated guideline for assessing discriminant validity. *Organizational Research Methods*, *25*(1), 6–14. https://doi.org/10.1177/1094428120968614

Workflow for discriminant validity analysis and reporting

Generalized Definition of Discriminant Validity

We present a definition that does not depend on a particular model and makes it explicit that discriminant validity is a feature of a measure instead of a construct:[2] *Two measures intended to measure distinct constructs have discriminant validity if the absolute value of the correlation between the measures after correcting for measurement error is low enough for the measures to be regarded as measuring distinct constructs.*

The objective of the workflow

1. Determine a level (no problem, three levels of problem) for each correlation pair

Table 12. Proposed Classification and Cutoffs.

Classification	CI_{CFA} (sys)	χ^2 (sys)
Severe problem	$1 \leq UL$	$\chi^2_1 - \chi^2_{org} < 3.84$
Moderate problem	$.9 \leq UL < 1$	Not "Marginal problem" AND $\chi^2_1 - \chi^2_{org} > 3.84$
Marginal problem	$.8 \leq UL < .9$	Not "No problem" AND$\chi^2_{.9} - \chi^2_{org} > 3.84$
No problem	$UL < .8$	$\rho_{CFA} < .8$ AND $\chi^2_{.8} - \chi^2_{org} > 3.84$

Note: ρ_{CFA} is the correlation obtained using CFA, UL is the 95% upper limit of ρ_{CFA} when $\rho_{CFA} > 0$, and the absolute value of the 95% lower limit of ρ_{CFA} when $\rho_{CFA} < 0$, χ^2_{org} is the chi-square value of the original model, and χ^2_c is the chi-square value of the comparison model where the focal correlation is fixed to c when $\rho_{CFA} > 0$ and $-c$ when $\rho_{CFA} < 0$.

2. Deal with any problems based on their level

Rönkkö, M., & Cho, E. (2022). An updated guideline for assessing discriminant validity. *Organizational Research Methods*, *25*(1), 6–14. https://doi.org/10.1177/1094428120968614

Two workflows for assessing discriminant validity correlations

CI_{CFA}(sys)

- Estimate a CFA model
- Compare the CI upper limits against the classification system cutoffs (lower limits for negative correlations)

Easier to apply and less likely to misused

Recommended alternative

χ^2 (sys)

- Estimate a CFA model
- Compare the correlation estimates against the classification system to determine intial level for each correlation
- Test each correlation against the upper limit of its current level using nested model χ^2 test

Slightly better statistical properties

Rönkkö, M., & Cho, E. (2022). An updated guideline for assessing discriminant validity. *Organizational Research Methods*, *25*(1), 6–14. https://doi.org/10.1177/1094428120968614

Generalized Definition of Discriminant Validity

We present a definition that does not depend on a particular model and makes it explicit that discriminant validity is a feature of a measure instead of a construct:[2] *Two measures intended to measure distinct constructs have discriminant validity if the absolute value of the correlation between the measures after correcting for measurement error is low enough for the measures to be regarded as measuring distinct constructs.*

Workflow CI_{CFA}(sys)

1. Determine what is "low enough"
   - .80 is a conservative starting point
   - Sometimes higher values needed (e.g. sex and gender identity)
   - Not applicable to continuum constructs
2. Estimate a CFA model and inspect confidence intervals

Table 12. Proposed Classification and Cutoffs.

Classification	CI_{CFA} (sys)	χ^2 (sys)
Severe problem	$1 \leq UL$	$\chi^2_1 - \chi^2_{org} < 3.84$
Moderate problem	$.9 \leq UL < 1$	Not "Marginal problem" AND $\chi^2_1 - \chi^2_{org} > 3.84$
Marginal problem	$.8 \leq UL < .9$	Not "No problem" AND $\chi^2_{.9} - \chi^2_{org} > 3.84$
No problem	$UL < .8$	$\rho_{CFA} < .8$ AND $\chi^2_{.8} - \chi^2_{org} > 3.84$

Note: ρ_{CFA} is the correlation obtained using CFA, UL is the 95% upper limit of ρ_{CFA} when $\rho_{CFA} > 0$, and the absolute value of the 95% lower limit of ρ_{CFA} when $\rho_{CFA} < 0$, χ^2_{org} is the chi-square value of the original model, and χ^2_c is the chi-square value of the comparison model where the focal correlation is fixed to c when $\rho_{CFA} > 0$ and $-c$ when $\rho_{CFA} < 0$.

Workflow χ^2 (sys)

1. Determine what is “low enough”
   - .80 is a conservative starting point
   - Sometimes higher values needed (e.g. sex and gender identity)
   - Not applicable to continuum constructs
2. Estimate a CFA model
   1. Inspect correlation estimate to determine a starting level
   2. Perform a nested model comparison test against a comparison model with correlation constrained to upper limit of the level
   3. If significant, the correlation is at the current level. If not, test against the next level

Table 12. Proposed Classification and Cutoffs.

Classification	CI_{CFA} (sys)	χ^2 (sys)
Severe problem	$1 \leq UL$	$\chi^2_1 - \chi^2_{org} < 3.84$
Moderate problem	$.9 \leq UL < 1$	Not “Marginal problem” AND $\chi^2_1 - \chi^2_{org} > 3.84$
Marginal problem	$.8 \leq UL < .9$	Not “No problem” AND$\chi^2_{.9} - \chi^2_{org} > 3.84$
No problem	$UL < .8$	$\rho_{CFA} < .8$ AND $\chi^2_{.8} - \chi^2_{org} > 3.84$

Note: ρ_{CFA} is the correlation obtained using CFA, UL is the 95% upper limit of ρ_{CFA} when $\rho_{CFA} > 0$, and the absolute value of the 95% lower limit of ρ_{CFA} when $\rho_{CFA} < 0$, χ^2_{org} is the chi-square value of the original model, and χ^2_c is the chi-square value of the comparison model where the focal correlation is fixed to c when $\rho_{CFA} > 0$ and $-c$ when $\rho_{CFA} < 0$.

What if we find problems?

Table 12. Proposed Classification and Cutoffs.

Classification	CI_{CFA} (sys)	χ^2 (sys)
Severe problem	$1 \leq UL$	$\chi^2_1 - \chi^2_{org} < 3.84$
Moderate problem	$.9 \leq UL < 1$	Not "Marginal problem" AND $\chi^2_1 - \chi^2_{org} > 3.84$
Marginal problem	$.8 \leq UL < .9$	Not "No problem" AND $\chi^2_{.9} - \chi^2_{org} > 3.84$
No problem	$UL < .8$	$\rho_{CFA} < .8$ AND $\chi^2_{.8} - \chi^2_{org} > 3.84$

Note: ρ_{CFA} is the correlation obtained using CFA, UL is the 95% upper limit of ρ_{CFA} when $\rho_{CFA} > 0$, and the absolute value of the 95% lower limit of ρ_{CFA} when $\rho_{CFA} < 0$, χ^2_{org} is the chi-square value of the original model, and χ^2_c is the chi-square value of the comparison model where the focal correlation is fixed to c when $\rho_{CFA} > 0$ and $-c$ when $\rho_{CFA} < 0$.

- If all correlations are in no problem class, no action is required
- For all problematic correlations, determine
  - The source of the problem
  - The magnitude of the problem

What actions different levels imply?

Table 12 shows the classification system we propose. We emphasize that these are guideline that can be adjusted case-by-case if warranted by theoretical understanding of the two constructs and measures, not strict rules that should always be followed. Based on our review, correlations below .8 were seldom considered problematic, and this is thus used as the cutoff for the first class, "No problem," which strictly speaking is not a proof of no problem, just no evidence of a problem. When correlations fall into this class, researchers can simply declare that they did not find any evidence of a discriminant validity problem. The next three steps are referred to as Marginal, Moderate, and Severe problems, respectively. The Severe problem is the most straightforward: two items or scales cannot be distinguished empirically, and researchers should rethink their concept definitions, measurement, or both. In empirical applications, the correlation level of .9 was nearly universally interpreted as a problem, and we therefore use this level as a cutoff between the Marginal and Moderate cases. In both cases, the high correlation should be acknowledged, and its possible cause should be discussed. In the Marginal case, the interpretation of the scales as representations of distinct constructs is probably safe. In the Moderate case, additional evidence from prior studies using the same constructs and/or measures should be checked before interpretation of the results to ensure that the high correlation is not a systematic problem with the constructs or scales.

Rönkkö, M., & Cho, E. (2022). An updated guideline for assessing discriminant validity. *Organizational Research Methods*, *25*(1), 6–14. https://doi.org/10.1177/1094428120968614

What to Do When Discriminant Validity Fails?

If problematically high correlations are observed, their sources must be identified. We propose a three-step process: First, suspect conceptual redundancy. We suggest starting by following the guidelines by J. A. Shaffer et al. (2016) and Podsakoff et al. (2016) for assessing the conceptual distinctiveness of the constructs (see also M. S. Krause, 2012). If two constructs are found to overlap conceptually, researchers should seriously consider dropping one of the constructs to avoid the confusion caused by using two different labels for the same concept or phenomenon (J. A. Shaffer et al., 2016).

Second, scrutinize the measurement model. An unexpectedly high correlation estimate can indicate a failure of model assumptions, as demonstrated by our results of misspecified models. Check the χ^2 test for an exact fit of the CFA model. If this test fails, diagnose the model with residuals and/or modification indices to understand the source of misspecification (Kline, 2011, chap. 8). If the model is modified based on these considerations, the wording of the items that led to these decisions should be explicitly reported, and how the item wordings justify the modifications should be explained to reduce the risk of data mining.

Third, collect different data. If conceptual overlap and measurement model issues have been ruled out, the discriminant validity problem can be reduced to a multicollinearity problem. For example, if one wants to study the effects of hair color and gender on intelligence but samples only blonde men and dark-haired women, hair color and gender are not empirically distinguishable, although they are both conceptually distinct and virtually uncorrelated in the broader population. This can occur either because of a systematic error in the sampling design or due to chance in small samples. If a systematic error can be ruled out, the most effective remedy is to collect more data. Alternatively, the data can be used as such, in which case large standard errors will indicate that little can be said about the relative effects of the two variables, or the two variables can be combined as an index (Wooldridge, 2013, pp. 94–98). If a researcher chooses to interpret results, he or she should clearly explain why the large correlation between the latent variables (e.g., >.9) is not a problem in the particular study.

Rönkkö, M., & Cho, E. (2022). An updated guideline for assessing discriminant validity. *Organizational Research Methods*, *25*(1), 6–14. https://doi.org/10.1177/1094428120968614

Summary of CI_{CFA}(sys) and χ^2 (sys)

- Choose context specific cutoffs if possible
- Classify each correlation as (example cutoff)
  - No problem (<.8)
  - Marginal problem (.8-.9)
  - Moderate problem (.9-1)
  - Severe problem (not different from 1)
- For all correlations that have problems, identify the sources
  - Conceptual overlap
  - Measurement problem
  - Sampling problem
- Explain the problems
  - Scale pairs with severe problems are not empirically distinct and cannot be used
  - Scale pairs with moderate problems can be used only if additional evidence from prior studies indicates this is not a systematic problem

How to Implement the Proposed Techniques

The proposed classification system should be applied with CI_{CFA}(cut) and χ^2(cut), and we propose that these workflows be referred to as CI_{CFA}(sys) and χ^2(sys), respectively. Both workflows start by estimating a CFA model that includes all scales that are evaluated for discriminant validity. Instead of using the default scale setting option to fix the first factor loadings to 1, scale the latent variables by fixing their variances to 1 (A in Figure 2); this should be explicitly reported in the article. The covariances between factors obtained in the latter way equal the correlations; alternatively, when using CI_{CFA}(sys), the standardized factor solution can be inspected. Next, inspect the upper limits (lower limits for negative correlations) of the 95% CIs of the estimated factor correlations and compare their values against the cutoffs in Table 12.[19]

Reporting

We also provide a few guidelines for improved reporting. First, researchers should clearly indicate what they are assessing when assessing discriminant validity by stating, for example, that "We addressed discriminant validity (whether two scales are empirically distinct)." Second, the correlation tables, which are ubiquitous in organizational research, are in most cases calculated with scale scores or other observed variables. However, most studies use only the lower triangle of the table, leaving the other half empty (*AMJ* 93.6%, *JAP* 83.1%). This practice is a waste of scarce resources, and we suggest that this space should be used for the latent correlation estimates, which serve as continuous discriminant validity evidence. Third, if nested model comparisons (e.g., $\chi^2(1)$) are used, researchers should explicitly report that the model was rescaled from the default option by stating, for example, "We used the χ^2 nested model comparison for assessing discriminant validity by comparing our CFA model against models that were more constrained, where all factor loadings were freely estimated, the factor variances were constrained to 1 and each factor correlation was constrained to 1 one at a time." These reporting practices should considerably reduce the ambiguity in the literature and prevent the common misapplication of the $\chi^2(1)$ test.

Rönkkö, M., & Cho, E. (2022). An updated guideline for assessing discriminant validity. *Organizational Research Methods*, *25*(1), 6–14. https://doi.org/10.1177/1094428120968614

Summary of discriminant and convergent validity assesment

	Convergent validity	Discriminant validity
Meaning	Do different measures measure the same thing	Do different measures measure different things
When needed	• When distinct measures are used (e.g. proxies) • Not if a correlation is used in a unidimensional reliability measure (e.g., alpha)	• There is a concern that measures intended to capture different things actually capture the same thing • Crrelations between scales are high (e.g. factor correlations >.50)
Statistical technique	Correlation between measures	Confirmatory factor analysis • CI of factor correlation recommended • Nested model χ^2 test can be used
Cutoffs	• .7 or more desirable • .5 or less should not be used	• No problem (<.8) • Marginal problem (.8-.9) • Moderate problem (.9-1) • Severe problem (not different from 1)
Failure cases	• Poor proxies	• Conceptual overlap • Measurement problem • Sampling problem