Scientific Research Publishing

Open Journal of Social Sciences, 2026, 14(11), 104-124
https://www.scirp.org/journal/jss
ISSN Online: 2327-5960
ISSN Print: 2327-5952

Structural Equation Modeling in Social Sciences: A Comparative Guide to Covariance-Based and Partial Least Squares Approaches

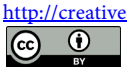
Nammuni Arachchige Nuwan Janaka Maduwansha

Department of Social Statistics, University of Sri Jayewardenepura, Nugegoda, Sri Lanka
Email: jnuwan66@sjp.ac.lk

How to cite this paper: Maduwansha, N. A. N. J. (2026). Structural Equation Modeling in Social Sciences: A Comparative Guide to Covariance-Based and Partial Least Squares Approaches. *Open Journal of Social Sciences, 14,* 104-124.
https://doi.org/10.4236/jss.2026.1411008

Received: September 25, 2026
Accepted: November 8, 2026
Published: November 11, 2026

Copyright © 2026 by author(s) and Scientific Research Publishing Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/



Open Access

Abstract

Structural equation modeling (SEM) is one of the most widely used multivariate techniques in the social and behavioral sciences, allowing researchers to test complex networks of relationships among observed and latent variables simultaneously, including direct, mediating, and moderating effects. Despite its popularity, researchers frequently struggle to distinguish between the two dominant estimation traditions, covariance-based SEM (CB-SEM) and partial least squares SEM (PLS-SEM), and to apply the correct evaluation criteria for each. This article provides a pedagogically oriented, comparative overview of SEM intended for social science researchers. It defines SEM and its measurement and structural components, and explains how these components extend beyond simple direct-effects models to accommodate mediation analysis, in which an intervening variable transmits or partially transmits the effect of a predictor on an outcome, and moderation analysis, in which the strength or direction of a relationship depends on a third, conditioning variable. The article then contrasts the theoretical assumptions and objectives of CB-SEM and PLS-SEM, including their differing suitability for testing indirect and conditional effects, and consolidates the recommended goodness-of-fit, normality, reliability, and validity criteria, together with their accepted threshold values, for each approach. It further outlines how mediating and moderating effects are estimated and assessed under each approach, sets out separate assessment criteria for reflective and formative measurement models, distinguishes in-sample predictive relevance from out-of-sample prediction, and cautions that directional paths estimated from observational data do not by themselves establish causal effects. By synthesizing guidance from foundational and contemporary methodological sources, the article aims to serve as a practical, citation-

supported reference for researchers designing, estimating, and reporting SEM-based research, including studies that model mediating and moderating relationships among latent constructs.

Keywords

Structural Equation Modeling, Covariance-Based SEM, Partial Least Squares SEM, Goodness-of-Fit, Validity, Reliability

1. Introduction

Structural equation modeling (SEM) has, in recent decades, become one of the primary data analysis techniques used by researchers across the social sciences, including sociology, psychology, education, management, and public policy (Hair et al., 2019b). Its growing dominance in applied social research reflects its unique capacity to examine multiple, interrelated dependence relationships simultaneously while explicitly accounting for measurement error in latent, unobserved constructs such as attitudes, satisfaction, trust, or social capital (Byrne, 2016; Kline, 2015). Unlike simpler techniques such as multiple regression or exploratory factor analysis considered in isolation, SEM integrates measurement theory and structural (path) theory into a single, testable statistical model (Jöreskog, 1969), which is a major reason it has displaced simpler multivariate techniques as the default confirmatory and predictive analysis method in journals across the social sciences.

Since Jöreskog's (1969) pioneering work on confirmatory maximum-likelihood factor analysis and Wold's (1982) development of partial least squares path modeling, SEM has evolved into a family of related but methodologically distinct techniques. Two estimation traditions dominate contemporary applied research: covariance-based SEM (CB-SEM), most commonly implemented in software such as AMOS, LISREL, and Mplus, and variance-based, or partial least squares, SEM (PLS-SEM), most commonly implemented in SmartPLS or WarpPLS (Hair et al., 2019b; Sarstedt et al., 2017). Although both approaches share the same conceptual foundation of relating latent constructs to their indicators and to one another, they differ substantially in their statistical objectives, distributional assumptions, sample size requirements, and, critically for applied researchers, the criteria used to judge whether an estimated model is acceptable (Hair et al., 2019b).

A recurring difficulty for graduate students and novice researchers is that the evaluation criteria developed for CB-SEM are not appropriate for PLS-SEM, and vice versa, yet these criteria are frequently confused or misapplied in student theses and even in published research (Hair et al., 2019b; Henseler et al., 2015). The purpose of this article is therefore twofold: first, to provide a clear, citation-supported explanation of what SEM is and how CB-SEM and PLS-SEM differ; and second, to consolidate, in tabular form, the goodness-of-fit, normality, and valid-

ity/reliability criteria recommended for each approach, together with their accepted threshold values. In addition, the article outlines how each approach estimates and assesses mediation and moderation, how formative measurement models should be evaluated, and what SEM results can and cannot establish about causal effects. The article is written primarily as a teaching and reference resource for social science researchers and students who are approaching SEM for the first time, or who need a single, well-referenced source to guide model evaluation and reporting.

2. What Is Structural Equation Modeling?

Structural equation modeling is a family of multivariate statistical techniques that combines elements of factor analysis and multiple regression analysis to estimate a network of relationships between measured (observed) variables and unmeasured (latent) constructs (Byrne, 2016; Hair et al., 2019a). A latent construct, such as job satisfaction, quality of life, or social trust, cannot be observed or measured directly; instead, it is inferred from a set of observed indicators (also called items or manifest variables) that are assumed to reflect, or in some cases form, the underlying construct (Kline, 2015).

A complete SEM specification generally consists of two interrelated sub-models (Hair et al., 2019a; Kline, 2015). The measurement model specifies the relationships between each latent construct and its observed indicators. It is conceptually equivalent to a confirmatory factor analysis (CFA) and allows the researcher to assess how well the chosen indicators represent their intended latent constructs before proceeding to test substantive hypotheses (Byrne, 2016). The structural model specifies the hypothesized directional or non-directional relationships among the latent constructs themselves, analogous to a system of simultaneous regression equations. Constructs that only receive paths from other constructs are termed endogenous, while constructs that only send paths to other constructs are termed exogenous (Kline, 2015).

Although the paths in a structural model are drawn as single-headed arrows and are commonly described as "effects," their direction is an assumption that the researcher specifies on theoretical grounds; it is not something that the estimation procedure discovers or tests. A model that fits the data well, or that yields statistically significant path coefficients, shows only that the data are consistent with the hypothesized structure, and equivalent models with reversed or rearranged paths will often reproduce the data equally well (MacCallum et al., 1993). This limitation is particularly important in the social sciences, where most SEM applications rely on observational, and frequently cross-sectional, survey data. In such studies, path coefficients estimate associations conditional on the model's assumptions; interpreting them as causal effects additionally requires that the cause precedes the effect, that relevant confounders have been measured and modeled, and that the hypothesized direction is correct. These conditions must be justified by the research design and by theory rather than by SEM itself (Antonakis et al., 2010;

Bollen & Pearl, 2013; Kline, 2015). Unless the design supports stronger claims, for example through random assignment, longitudinal panel data, or valid instrumental variables, researchers should describe directional paths in associational terms (e.g., "is positively associated with" or "predicts") rather than as evidence that one construct "causes" or "leads to" another.

Because SEM estimates the measurement and structural components simultaneously, it explicitly separates random measurement error from the true relationships among constructs, which is a key advantage over path analysis performed on composite or observed scores alone (Hair et al., 2019a). This property makes SEM particularly well suited to social science research, where constructs of interest are almost always latent and are measured with survey items subject to both random and systematic error (Kline, 2015).

Two broad estimation traditions have emerged for fitting SEM models to data. Covariance-based SEM (CB-SEM), rooted in the work of Jöreskog (1969) and implemented through maximum-likelihood estimation, seeks parameter estimates that minimize the discrepancy between the observed sample covariance matrix and the covariance matrix implied by the specified model (Byrne, 2016). Partial least squares SEM (PLS-SEM), rooted in Wold's (1982) soft-modeling tradition and popularized by Chin (1998) and Hair et al. (2019b), instead uses an iterative, component-based (ordinary least squares) algorithm that maximizes the explained variance of endogenous constructs.

3. Difference between CB-SEM and PLS-SEM

Although CB-SEM and PLS-SEM are both referred to as "structural equation modeling," they rest on different statistical logics and are suited to different research objectives (Hair et al., 2019b). CB-SEM is fundamentally a covariance-fitting, theory-confirmation technique. Because it evaluates how closely the model-implied covariance matrix reproduces the observed covariance matrix, it produces a set of global goodness-of-fit statistics that indicate whether the overall model is consistent with the data (Hu & Bentler, 1999). CB-SEM assumes multivariate normality of the observed indicators, generally requires a comparatively larger sample size for stable maximum-likelihood estimation, and is best suited to well-established theoretical models in which the researcher wishes to confirm or disconfirm a specific covariance structure (Byrne, 2016; Hair et al., 2019a).

PLS-SEM, in contrast, is a prediction-oriented, variance-maximizing technique. Because it does not rely on a common-factor model or on covariance reproduction, PLS-SEM does not produce a single chi-square-based test of overall model fit; instead, model quality is assessed through a series of measurement and structural model criteria focused on reliability, validity, and predictive power (Hair et al., 2019b). PLS-SEM makes minimal distributional assumptions, performs well with smaller sample sizes and more complex models involving many constructs and indicators, and can accommodate both reflective and formative measurement models, whereas CB-SEM handles formative constructs only with

considerable additional constraints (Chin, 1998; Hair et al., 2019b). The main differences between the two approaches are summarized in Table 1.

Table 1. Comparison of covariance-based SEM and partial least squares SEM.

Criterion	CB-SEM	PLS-SEM
Statistical objective	Theory testing/confirmation; reproduces the observed covariance matrix	Theory building/prediction; maximizes explained variance (R^2) of endogenous constructs
Underlying model	Common factor model	Composite/component-based model
Estimation method	Maximum likelihood (typically)	Iterative ordinary least squares (partial least squares algorithm)
Distributional assumption	Assumes multivariate normality	Distribution-free (non-parametric); robust to non-normal data
Typical minimum sample size	Larger samples generally required for stable estimation (often > 200)	Performs well with smaller samples; handles complex models with modest samples
Model complexity	Estimation difficulty increases sharply with many constructs/indicators	Handles complex models with many constructs and indicators efficiently
Indicator type	Primarily reflective; formative indicators require special handling	Readily handles both reflective and formative indicators
Global model fit	Provides chi-square and global fit indices (CFI, RMSEA, RMR, etc.)	No single global fit test; approximate fit measures have been proposed
Typical software	AMOS, LISREL, Mplus, EQS	SmartPLS, WarpPLS, ADANCO, R packages
Best suited for	Confirming/testing well-developed theory; strict hypothesis testing	Exploratory research, prediction, complex models, formative constructs, early-stage theory

Note. Adapted from guidance in Chin (1998), Hair et al. (2019b), Jöreskog (1969), and Wold (1982).

In practice, the choice between CB-SEM and PLS-SEM should be guided by the research objective rather than by convenience or software familiarity. Hair et al. (2019b) recommend CB-SEM when the research goal is theory confirmation and the model is grounded in well-developed theory, and PLS-SEM when the goal is theory development, prediction, or exploratory analysis, especially under conditions of small sample sizes, non-normal data, formative measurement, or highly complex models. Sarstedt et al. (2017) similarly emphasize that PLS-SEM should not be viewed merely as a "fallback" option for CB-SEM but as a distinct methodological choice suited to prediction-oriented research questions.

4. Evaluating CB-SEM Models

Because CB-SEM produces a model-implied covariance matrix that can be directly compared with the observed covariance matrix, its evaluation is organized around three broad categories of statistics: goodness-of-fit indices, which assess how well the overall model reproduces the observed data; normality diagnostics, which check whether the assumptions of maximum-likelihood estimation are met; and reliability/validity measures, which assess the quality of the measurement model (Byrne, 2016; Hair et al., 2019a; Kline, 2015).

4.1. Goodness-of-Fit Measures in CB-SEM

AMOS, the most widely used CB-SEM software package, presents its Model Fit Summary in ten output blocks: CMIN (chi-square statistics), RMR/GFI, Baseline Comparisons, Parsimony-Adjusted Measures, NCP, FMIN, RMSEA, information criteria (AIC-family), ECVI, and HOELTER (Hooper et al., 2008). These indices are conventionally classified into absolute fit indices, which assess how well the specified model reproduces the observed covariance matrix without reference to a baseline model; incremental (comparative) fit indices, which compare the fit of the specified model to a null/independence model; parsimony fit indices, which adjust fit statistics for model complexity; and information-theoretic/sample-adequacy indices, which are used for comparing competing models or checking sample size sufficiency (Bagozzi & Yi, 1988; Hu & Bentler, 1999). **Table 2** lists the indices produced by AMOS's default Model Fit Summary, together with their recommended threshold values.

Table 2. Goodness-of-fit indices reported by AMOS, with recommended values.

Index	Recommended value	Source
Chi-square (CMIN), df, p	Non-significant $p > 0.05$ is ideal but rarely attained with large samples; reported for transparency alongside other indices	Jöreskog (1969); Hooper et al. (2008)
Normed chi-square (CMIN/DF)	≤3.00 (good); ≤5.00 (acceptable)	Kline (2015); Hooper et al. (2008)
Root Mean Square Residual (RMR)	No universal cutoff: RMR is expressed in the units of the observed variables, so its size depends on their measurement scales (0 = perfect fit; smaller is better relative to those scales). For a scale-free criterion, use the standardized RMR (SRMR ≤ 0.08), which AMOS reports only through its Standardized RMR plugin.	Hooper et al. (2008); Kline (2015)
Goodness-of-Fit Index (GFI)	≥0.95 (good); ≥0.90 (acceptable)	Jöreskog & Sörbom (1984); Hooper et al. (2008)
Adjusted GFI (AGFI)	≥0.90	Jöreskog & Sörbom (1984)
Parsimony GFI (PGFI)	>0.50 (used comparatively across competing models)	Byrne (2016)
Normed Fit Index (NFI)	≥0.95 (good); ≥0.90 (acceptable)	Bentler & Bonett (1980)
Relative Fit Index (RFI)	≥0.90 (values closer to 1 preferred)	Bollen (1989); Hooper et al. (2008)
Incremental Fit Index (IFI)	≥0.95 (good); ≥0.90 (acceptable)	Bollen (1989)
Tucker-Lewis Index (TLI)	≥0.95 (good); ≥0.90 (acceptable)	Hu & Bentler (1999)
Comparative Fit Index (CFI)	≥0.95 (good); ≥0.90 (acceptable)	Hu & Bentler (1999); Bentler (1990)
Parsimony Ratio (PRATIO)	Reported for context; used to compute PNFI/PCFI rather than judged alone	Byrne (2016)
Parsimony NFI (PNFI)	>0.50	Byrne (2016)
Parsimony CFI (PCFI)	>0.50	Byrne (2016)

Continued

Non-Centrality Parameter (NCP), with 90% CI	Closer to 0 is better; used mainly to compare nested/competing models	Hooper et al. (2008)
Minimum discrepancy function (FMIN), with F0	Closer to 0 is better	Hooper et al. (2008)
Root Mean Square Error of Approximation (RMSEA)	≤0.06 (good); ≤0.08 (acceptable)	Hu & Bentler (1999); Steiger (1990)
RMSEA 90% CI (LO90, HI90) and PCLOSE	HI90 upper bound < 0.10; PCLOSE (test of close fit) > 0.05	Browne & Cudeck (1993)
Akaike Information Criterion (AIC)	Lower value indicates better relative fit among competing, non-nested models	Akaike (1987)
Browne-Cudeck Criterion (BCC)	Lower value preferred; adjusts AIC for sample size	Browne & Cudeck (1989)
Bayesian Information Criterion (BIC)	Lower value preferred; imposes a stronger complexity penalty than AIC	Byrne (2016)
Consistent AIC (CAIC)	Lower value preferred	Bozdogan (1987)
Expected Cross-Validation Index (ECVI)	Lower value preferred, relative to the saturated and independence models' ECVI	Browne & Cudeck (1989)
Hoelter's critical N (.05 and .01)	≥200 indicates the sample is adequate for the model at the given significance level	Hoelter (1983)

It is important to note that Hu and Bentler's (1999) widely cited cutoff values, while highly influential, were derived under specific simulation conditions and should be interpreted as general guidelines rather than rigid pass/fail rules (Kline, 2015). In practice, researchers rarely report all of AMOS's fit indices in a single manuscript; rather, they select a small, complementary combination, typically CMIN/DF, one incremental index (e.g., CFI or TLI), and RMSEA with its confidence interval, often alongside RMR, while the remaining indices (e.g., AIC, ECVI, Hoelter's N) are reserved for comparing alternative or nested models rather than for judging a single model in isolation (Hooper et al., 2008; Hu & Bentler, 1999; Kline, 2015). Fit criteria should ideally be established a priori rather than selected post hoc.

4.2. Normality Measures in CB-SEM

Because CB-SEM's default maximum-likelihood estimator assumes multivariate normality of the observed indicators, researchers are expected to assess normality before interpreting chi-square-based fit statistics and standard errors (Kline, 2015). **Table 3** summarizes the commonly used univariate and multivariate normality diagnostics.

Mardia's (1970) multivariate kurtosis coefficient is interpreted relative to its expected value under multivariate normality, $p(p+2)$, where p denotes the number of observed indicator variables included in the model; for example, with $p = 10$ indicators, the expected kurtosis coefficient under normality is $10(10 + 2) = 120$. Software such as AMOS reports the observed coefficient alongside a normalized estimate (critical ratio) that tests the statistical significance of the deviation be-

tween the observed value and this $p(p + 2)$ expectation (Mardia, 1970; Byrne, 2016). Because this critical ratio is interpreted as an approximately standard normal (z) statistic, an absolute value greater than 1.96 already indicates a statistically significant departure from multivariate normality at the .05 level, and a value near 5.00 represents a highly significant departure. Values greater than 5.00 are conventionally treated as indicating data that are non-normally distributed (Byrne, 2016). The 5.00 value should therefore be understood as a rough marker of severe non-normality, below which researchers sometimes judge maximum-likelihood estimation to be tolerably robust, and not as evidence that the data are multivariate normal. Because the test becomes very sensitive in large samples, its result should be read alongside the size of the kurtosis coefficient relative to $p(p + 2)$ and the univariate diagnostics, and robust estimation is advisable whenever non-normality is in doubt (Kline, 2015).

Table 3. Recommended normality diagnostics for CB-SEM.

Measure	Recommended value	Source
Univariate skewness	Absolute value < 3 (values > 3 indicate extreme skewness); a stricter guideline of <2 is also used	Kline (2015); West et al. (1995)
Univariate kurtosis	Absolute value < 10 (values > 20 indicate severe kurtosis); a stricter guideline of <7 is also used	Kline (2015); West et al. (1995)
Mardia's multivariate skewness coefficient	Value close to 0 indicates multivariate normality; evaluated via its associated significance test	Mardia (1970); Byrne (2016)
Mardia's multivariate kurtosis coefficient	Under multivariate normality, the expected value of the coefficient is $p(p + 2)$, where p is the number of observed indicator variables in the model	Mardia (1970); Byrne (2016)
Mardia's multivariate kurtosis (normalized/critical-ratio estimate)	Interpreted as a z statistic testing the null hypothesis of multivariate normality: |c.r.| > 1.96 indicates a significant departure at $p < 0.05$; values > 5.00 are conventionally treated as indicating non-normal data. A value below 5.00 is not evidence of normality; it only falls below this conventional severity heuristic	Mardia (1970); Byrne (2016)

When indicators substantially violate multivariate normality, researchers commonly apply robust or scaled estimation methods (e.g., Satorra-Bentler corrections), bootstrap-based standard errors, or consider whether a distribution-free approach such as PLS-SEM is more appropriate for the data at hand (Hair et al., 2019a; Kline, 2015).

4.3. Reliability and Validity Measures in CB-SEM

Prior to interpreting the structural model, researchers using CB-SEM must establish that the measurement model demonstrates adequate reliability and validity (Hair et al., 2019a). **Table 4** presents the most commonly reported criteria.

Convergent validity is established when indicators of the same construct share a high proportion of variance, typically evidenced by AVE values of at least .50, meaning the construct explains more than half of the variance in its indicators (Fornell & Larcker, 1981). Discriminant validity, which confirms that a construct is empirically distinct from other constructs in the model, is tradi-

tionally assessed using the Fornell-Larcker (Fornell & Larcker, 1981) criterion: the square root of each construct's AVE should exceed its correlations with all other constructs in the model.

Table 4. Recommended reliability and validity criteria for CB-SEM.

Measure	Type	Recommended value	Source
Cronbach's alpha (α)	Internal consistency reliability	≥0.70	Nunnally & Bernstein (1994)
Composite/construct reliability (CR)	Internal consistency reliability	≥0.70	Fornell & Larcker (1981); Hair et al. (2019a)
Standardized factor loadings	Indicator (item) reliability	≥0.50, ideally ≥ 0.70	Hair et al. (2019a)
Average Variance Extracted (AVE)	Convergent validity	≥0.50	Fornell & Larcker (1981)
Fornell-Larcker criterion (√AVE vs. inter-construct correlations)	Discriminant validity	√AVE for each construct exceeds its correlations with all other constructs	Fornell & Larcker (1981)

5. Evaluating PLS-SEM Models

Because PLS-SEM does not rely on covariance reproduction, it is evaluated using a systematic two-step process that first assesses the measurement (outer) model and then the structural (inner) model, rather than a single global fit statistic (Hair et al., 2019b). This section summarizes the criteria recommended for each step.

5.1. Reliability and Validity in PLS-SEM: Reflective Measurement Models

Table 5 summarizes the criteria recommended for assessing the reliability and validity of reflective measurement models in PLS-SEM. Formative measurement models require a different set of criteria, which are discussed in Section 5.2.

Table 5. Recommended reflective measurement model criteria for PLS-SEM.

Measure	Type	Recommended value	Source
Indicator (outer) loadings	Indicator reliability	≥0.7 (indicator explains > 50% of its variance); loadings of 0.40 - 0.70 retained only if removal does not improve AVE/CR	Hair et al. (2019b)
Cronbach's alpha (α)	Internal consistency reliability	≥0.70 (0.60 - 0.70 deemed acceptable in exploratory research)	Hair et al. (2019b); Nunnally & Bernstein (1994)
rho_A (Dijkstra-Henseler reliability coefficient)	Internal consistency reliability	≥0.70	Dijkstra & Henseler (2015)
Composite reliability (CR, ρc)	Internal consistency reliability	≥0.70 (0.60 - 0.70 acceptable in exploratory research); values > 0.95 may indicate item redundancy	Hair et al. (2019b)

Continued

Average Variance Extracted (AVE)	Convergent validity	≥0.50	Fornell & Larcker (1981); Hair et al. (2019b)
Fornell-Larcker criterion	Discriminant validity	√AVE for each construct exceeds its correlations with other constructs	Fornell & Larcker (1981)
Heterotrait-Monotrait ratio (HTMT)	Discriminant validity	<0.85 (strict); <0.90 acceptable for conceptually similar constructs	Henseler et al. (2015)

Hair et al. (2019b) note that the Fornell-Larcker criterion, while historically dominant, performs poorly at detecting discriminant validity problems under certain conditions, and recommend supplementing or replacing it with the heterotrait-monotrait ratio of correlations (HTMT), which compares the average correlations of indicators across constructs measuring different constructs to the average correlations of indicators within the same construct (Henseler et al., 2015). An HTMT value below 0.85 provides strong evidence of discriminant validity, while values below 0.90 are considered acceptable when the constructs being compared are conceptually more similar to one another (Henseler et al., 2015).

5.2. Assessing Formative Measurement Models in PLS-SEM

In a formative measurement model, the indicators are assumed to jointly define, or form, the construct rather than to reflect it, as when socioeconomic status is formed from income, education, and occupational prestige. The choice between reflective and formative specification should be made on theoretical grounds before estimation, because misspecifying a formative construct as reflective, or vice versa, can substantially bias structural estimates (Jarvis et al., 2003). Because formative indicators need not be correlated with one another and are not interchangeable, the reflective criteria in **Table 5**, namely internal consistency reliability (α, rho_A, and CR), AVE, the Fornell-Larcker criterion, and HTMT, are not meaningful for formative constructs and should not be reported for them (Diamantopoulos & Winklhofer, 2001; Hair et al., 2019b). Instead, Hair et al. (2019b) recommend assessing formative measurement models in three steps, covering convergent validity, indicator collinearity, and the statistical significance and relevance of the indicator weights, all of which presuppose that content validity was established when the indicators were selected. **Table 6** summarizes these criteria.

Removing a formative indicator changes the conceptual meaning of the construct, so indicators should not be deleted on statistical grounds alone (Cenfetelli & Bassellier, 2009; Diamantopoulos & Winklhofer, 2001). High indicator collinearity, by contrast, inflates the standard errors of the weights and can produce counterintuitive weight signs; it is usually addressed by merging or removing conceptually redundant indicators. Formative constructs can also be specified in CB-SEM, but identification normally requires the formative construct to emit paths to at least two reflectively measured variables, as in multiple indicators, multiple causes (MIMIC) specification. This additional complexity is one reason why PLS-SEM is often preferred when a model contains several formative constructs

(Diamantopoulos & Winklhofer, 2001; Jarvis et al., 2003).

Table 6. Recommended criteria for formative measurement models in PLS-SEM.

Measure	Type	Recommended value	Source
Content validity	Domain coverage	Indicators jointly capture all major facets of the construct's domain, established from theory, literature, and expert judgment before data collection	Diamantopoulos & Winklhofer (2001)
Redundancy analysis (formative construct → alternative reflective or global single-item measure of the same construct)	Convergent validity	Path coefficient ≥ 0.70 (the formative construct explains at least about 50% of the variance of the alternative measure)	Chin (1998); Cheah et al. (2018); Hair, Risher et al. (2019b)
Indicator variance inflation factor (VIF)	Indicator collinearity	<5.00 (ideally < 3); higher values signal critical collinearity among the formative indicators	Hair, Risher et al. (2019b); Diamantopoulos & Siguaw (2006)
Outer weights	Relative indicator contribution	Significant, i.e., bootstrap 95% confidence interval excludes zero (typically 5000 subsamples)	Hair, Risher et al. (2019b)
Outer loadings of formative indicators	Absolute indicator contribution	If a weight is non-significant, retain the indicator when its loading is ≥0.50 or significant; consider removal only when the loading is also low and non-significant and content validity is not compromised	Cenfetelli & Bassellier (2009); Hair, Risher et al. (2019b)
α, rho_A, CR, AVE, Fornell-Larcker, HTMT	Not applicable	Not meaningful for formative constructs and should not be reported for them	Diamantopoulos & Winklhofer (2001); Hair et al. (2019b)

5.3. Goodness-of-Fit (Model Fit) Measures in SmartPLS

For much of its history, PLS-SEM was considered a method without a model fit test, since it does not minimize a discrepancy function between an observed and model-implied covariance matrix (Henseler et al., 2014). In response to demand for fit diagnostics comparable to CB-SEM, SmartPLS now generates a dedicated "Model Fit" report alongside the standard PLS algorithm and PLSc (consistent PLS) results, comprising six measures: SRMR, the unweighted least squares discrepancy (d_ULS), the geodesic discrepancy (d_G), Chi-square, the Normed Fit Index (NFI), and, for composite-consistent (PLSc) estimation, RMS_theta (Dijkstra & Henseler, 2015; Henseler et al., 2014). Each measure is reported for both the saturated model, in which all constructs are allowed to correlate freely, and the estimated model, which imposes the researcher's hypothesized structural paths. **Table 7** summarizes these six SmartPLS model fit indices.

Hair et al. (2019b) and Henseler et al. (2014) caution that these fit measures were adapted from the CB-SEM tradition and remain more contested for PLS-SEM: because PLS-SEM optimizes prediction rather than covariance reproduc-

tion, a low SRMR or a passed d_ULS/d_G exact-fit test does not guarantee a theoretically meaningful or predictively useful model, and a comparatively higher value does not necessarily invalidate one. For this reason, SmartPLS model fit indices are generally reported as a complement to, rather than a substitute for, the measurement-model criteria in **Tables 5-6** and the prediction-oriented structural criteria in **Table 8** (Hair et al., 2019b).

Table 7. Goodness-of-fit (Model Fit) indices reported by SmartPLS.

Index	What it assesses	Recommended value	Source
Standardized Root Mean Square Residual (SRMR)	Average discrepancy between the observed and model-implied correlation matrix (saturated and estimated models)	<0.08 (some sources tolerate < 0.10); reported for both saturated and estimated models	Hu & Bentler (1999); Henseler et al. (2014); Hair et al. (2019b)
Unweighted L2 distance (d_ULS)	Squared Euclidean distance between the observed and model-implied correlation matrix	Exact-fit test: model supported if the original d_ULS value falls below the 95% (HI95) or 99% (HI99) bootstrap quantile	Dijkstra & Henseler (2015); Henseler et al. (2014)
Geodesic distance (d_G)	Geodesic discrepancy between the observed and model-implied correlation matrix	Exact-fit test: model supported if the original d_G value falls below the 95% (HI95) or 99% (HI99) bootstrap quantile	Dijkstra & Henseler (2015); Henseler et al. (2014)
Chi-square (χ^2)	Discrepancy statistic reported for reference, analogous to the CB-SEM chi-square	No fixed cutoff in PLS-SEM; lower values indicate closer correspondence between matrices; interpreted cautiously	Henseler et al. (2014)
Normed Fit Index (NFI)	Proportional improvement in fit relative to a null/independence model	≥0.90 (values closer to 1 preferred); reported for saturated and estimated models	Bentler & Bonett (1980); Henseler et al. (2014)
RMS_theta (composite/PLSc models only)	Degree of correlation among the outer residuals of a composite/reflective measurement model	<0.12 indicates a well-fitting measurement model	Henseler et al. (2014)

5.4. Explanatory and Predictive Power in PLS-SEM

Beyond the model fit indices above, structural model quality in PLS-SEM is assessed using a set of complementary, prediction-oriented criteria addressing collinearity, explanatory power, effect size, predictive relevance, out-of-sample predictive power, and path significance (Hair et al., 2019b). **Table 8** summarizes these criteria.

Table 8. Recommended structural model criteria for PLS-SEM.

Measure	Purpose	Recommended value	Source
Variance Inflation Factor (VIF)	Collinearity diagnostic among predictor constructs	<5.00 (ideally < 3)	Hair et al. (2019b); Diamantopoulos & Siguaw (2006)
Coefficient of determination (R^2)	Explanatory/in-sample predictive power of endogenous constructs	0.75 = substantial; 0.50 = moderate; 0.25 = weak	Hair et al. (2019b); cf. Chin (1998): 0.67 = substantial; 0.33 = moderate; 0.19 = weak

Continued

Effect size (f^2)	Contribution of each exogenous construct to the R^2 of an endogenous construct	0.02 = small; 0.15 = medium; 0.35 = large	Cohen (1988)
Predictive relevance (Q^2), via blindfolding	Predictive relevance assessed by omitting and imputing individual data points within the estimation sample; blends in-sample explanatory and predictive power and is not a test on held-out cases	$Q^2 > 0$ indicates predictive relevance	Stone (1974); Geisser (1974); Hair et al. (2019b)
PLSpredict ($Q^2_{predict}$, RMSE/MAE)	Out-of-sample predictive power, assessed on held-out observations via k-fold cross-validation (typically k = 10)	$Q^2_{predict} > 0$; RMSE (or MAE) lower than the linear model (LM) benchmark for all, the majority, a minority, or none of the indicators = high, medium, low, or no predictive power	Shmueli et al. (2016, 2019)
Path coefficients (β) and significance	Strength/significance of hypothesized structural relationships	Significant at $p < 0.05$ (two-tailed), assessed via bootstrapping (typically 5000 subsamples)	Hair et al. (2019b)

Because PLS-SEM parameter estimates are not based on a distributional assumption, significance testing of path coefficients and other parameters is conducted using nonparametric bootstrapping, typically with 5000 subsamples, rather than the asymptotic standard errors used in CB-SEM (Hair et al., 2019b). The Q^2 value obtained through blindfolding should not be described as a measure of out-of-sample prediction. Blindfolding omits individual data points, imputes them, and re-estimates the model using the remaining data from the same sample, so the cases whose values are predicted also contribute to estimating the model. Q^2 therefore blends in-sample explanatory power with predictive relevance rather than demonstrating how well the model predicts new observations (Hair et al., 2019b; Shmueli et al., 2016). Genuine out-of-sample prediction requires predictions for observations that were held out of model estimation. The PLSpredict procedure achieves this through k-fold cross-validation: the sample is split into k subsets (typically 10), the model is estimated on k − 1 subsets, and the indicators of the endogenous constructs are predicted for the held-out subset, rotating until every case has been predicted once (Shmueli et al., 2016, 2019). The results are interpreted against two benchmarks. $Q^2_{predict}$ values above zero indicate that the model's predictions outperform the naive benchmark of the training-sample indicator means, and the root mean squared error (RMSE), or the mean absolute error (MAE) when prediction errors are highly non-symmetric, is compared with that of a linear regression model (LM) benchmark. The model has high, medium, low, or no predictive power when it produces lower prediction errors than the LM benchmark for all, the majority, a minority, or none of the indicators, respectively

(Shmueli et al., 2019).

6. Mediation and Moderation in CB-SEM and PLS-SEM

Many social science models go beyond direct effects to ask how an effect is transmitted (mediation) or under what conditions it holds (moderation). Both CB-SEM and PLS-SEM can estimate such models, but they differ in how indirect and conditional effects are specified, estimated, and assessed. Because mediation describes a causal process, the caution in Section 2 about the causal interpretation of directional paths applies with particular force; in cross-sectional data, estimates of indirect effects can be substantially biased relative to the longitudinal processes they are meant to represent (Maxwell & Cole, 2007).

6.1. Mediation Analysis

In a simple mediation model, the indirect effect of a predictor X on an outcome Y through a mediator M is the product of the X → M path (a) and the M → Y path (b); the direct effect (c') is the X → Y path controlling for M, and the total effect is the sum of the direct and indirect effects. Contemporary guidance has moved away from Baron and Kenny's (1986) causal-steps procedure, which does not test the indirect effect itself, and from the Sobel test, which assumes that the product $a \times b$ is normally distributed, an assumption that rarely holds. Instead, the indirect effect is tested directly using bootstrap confidence intervals: mediation is supported when the 95% confidence interval for $a \times b$ excludes zero, and a significant total effect of X on Y is not a prerequisite (Preacher & Hayes, 2008; Zhao et al., 2010). Following Zhao et al. (2010), results can then be classified as complementary or competitive partial mediation (direct and indirect effects both significant, with the same or opposite signs), indirect-only (full) mediation (only the indirect effect significant), direct-only non-mediation, or no-effect non-mediation.

In CB-SEM, mediators are modeled as latent constructs within the full structural model, which corrects the a and b paths for measurement error and so avoids the attenuation that affects regression-based mediation analysis with composite scores (Kline, 2015). Model fit is evaluated as usual with the indices in **Table 2**, and alternative specifications, such as a model in which the direct path is constrained to zero, can be compared with a chi-square difference test. In AMOS, total, direct, and indirect effects are requested in the output and tested by enabling bootstrapping with percentile or bias-corrected confidence intervals; specific indirect effects in models with several mediators require user-defined estimands (Byrne, 2016). In PLS-SEM, the same logic applies without a global fit assessment: specific and total indirect effects are computed from the path estimates and tested with the same bootstrapping procedure (typically 5000 subsamples) used for the other structural paths, and the Zhao et al. (2010) typology is used to characterize the type of mediation (Hair et al., 2019b; Nitzl et al., 2016). In both approaches, the measurement models of all constructs involved, including the mediators, should meet the criteria in Sections 4.3 or 5.1 - 5.2 before indirect effects are interpreted.

6.2. Moderation Analysis

Moderation occurs when the strength or direction of the relationship between X and Y depends on a third variable, the moderator W. When the moderator is categorical (e.g., gender or country), moderation is commonly examined with multigroup analysis. In CB-SEM, multigroup analysis first establishes measurement invariance across groups, testing configural, metric (equal loadings), and, where latent means are compared, scalar (equal intercepts) invariance, and then tests whether a structural path differs between groups by comparing a model in which that path is constrained to be equal across groups with an unconstrained model using a chi-square difference test (Byrne, 2016). In PLS-SEM, the analogous procedure first assesses invariance using the measurement invariance of composites (MICOM) procedure and then compares group-specific path coefficients with permutation or bootstrap-based multigroup tests (Henseler et al., 2016).

When the moderator is continuous, an interaction term between the latent predictor and the latent moderator is added to the model. In CB-SEM, latent interactions are estimated either with product-indicator approaches, in which products of the (mean-centered) indicators of the two constructs serve as indicators of the latent interaction (Kenny & Judd, 1984; Marsh et al., 2004), or with distribution-analytic methods such as latent moderated structural equations (LMS), which account for the non-normality that interaction terms create (Klein & Moosbrugger, 2000). LMS is available in Mplus but not in AMOS, where researchers typically rely on product indicators, on interactions between observed composite scores (which do not correct for measurement error), or on multigroup analysis. Because LMS does not provide the conventional fit indices listed in **Table 2**, an interaction model estimated in this way is usually evaluated by comparing it with the corresponding main-effects model using a likelihood-ratio test. In PLS-SEM, interaction terms are created with the product-indicator, orthogonalizing, or two-stage approach; the two-stage approach, which forms the interaction term from the latent variable scores of a main-effects model, is generally recommended and is the appropriate choice when the predictor or the moderator is formatively measured (Becker et al., 2018; Henseler & Chin, 2010).

In both approaches, a moderating effect is supported when the path from the interaction term to Y is significant, assessed with critical ratios (or likelihood-ratio tests) in CB-SEM and with bootstrap confidence intervals in PLS-SEM. Because interaction effects in social science data are typically small, the f^2 of an interaction term should not be judged against Cohen's (1988) conventions for main effects (Aguinis et al., 2005). A significant interaction should be probed and interpreted, for example with simple-slope plots of the X → Y relationship at low, average, and high levels of the moderator. When a moderator is hypothesized to alter an indirect effect (moderated mediation), researchers should report the conditional indirect effects at selected values of the moderator together with a bootstrap confidence interval for the index of moderated mediation (Hayes, 2015).

7. Practical Guidance for Researchers

For researchers beginning to work with SEM, four practical recommendations follow from the material reviewed above. First, the choice between CB-SEM and PLS-SEM should be made before data collection, based on the research objective (theory confirmation versus prediction/theory development), the anticipated sample size, the expected distribution of the data, and whether the measurement model includes formative indicators (Hair et al., 2019b; Sarstedt et al., 2017). Switching approaches after the fact simply to obtain more favorable results is a questionable research practice and should be avoided (Hair et al., 2019b).

Second, model evaluation should always proceed in the correct order: the measurement model (reliability and validity) must be assessed and confirmed as adequate before the structural model, that is, the hypothesized relationships, is interpreted, regardless of whether CB-SEM or PLS-SEM is used (Hair et al., 2019b). Reporting structural path coefficients from a measurement model with poor reliability or validity produces results that cannot be meaningfully interpreted.

Third, researchers should resist the temptation to rely on a single statistic to justify model adequacy. Methodologists consistently recommend reporting a combination of indices (for example, RMSEA together with CFI and RMR in CB-SEM, or AVE, HTMT, and Q^2 together in PLS-SEM) and interpreting recommended threshold values as general guidelines rather than rigid pass/fail rules (Hu & Bentler, 1999; Kline, 2015; Hair et al., 2019b). Sample size, model complexity, and the number of indicators per construct can all influence fit statistics, and thresholds developed under one set of simulation conditions do not automatically generalize to every applied context (Kline, 2015). Ultimately, sound SEM practice combines statistical criteria with careful theoretical justification of the specified model (Byrne, 2016).

Fourth, researchers should match the strength of their conclusions to the strength of their research design. As noted in Section 2, a well-fitting model and significant directional paths do not by themselves establish that one construct causes another, particularly in observational and cross-sectional studies (Antonakis et al., 2010; Bollen & Pearl, 2013). Hypotheses may be derived from theory and tested as directional relationships, but the discussion of results should acknowledge plausible equivalent models, reverse causation, and omitted confounders, and should reserve causal language for designs, such as experiments or longitudinal panel studies, that support it.

8. A Note on Software and Reporting

Researchers should also be aware that the choice of software often follows directly from the choice of estimation approach. CB-SEM is most commonly estimated using IBM SPSS Amos, LISREL, Mplus, or the lavaan package in R; AMOS in particular produces the full ten-block Model Fit Summary summarized in **Table 2**, from CMIN and RMR/GFI through Hoelter's critical N, although SRMR, which

Table 2 also mentions, must be requested separately through the Standardized RMR plugin (Byrne, 2016; Hooper et al., 2008). PLS-SEM is most commonly estimated using SmartPLS, WarpPLS, ADANCO, or the seminr and plspm packages in R; SmartPLS in particular produces the bootstrapped measurement-model criteria in Tables 5-6, the SRMR/d_ULS/d_G/Chi-square/NFI/RMS_theta model fit report in Table 7, and the structural-model criteria in Table 8 (Hair et al., 2019b; Henseler et al., 2014). Because the two families of software report different default statistics, researchers should confirm, before beginning analysis, that the criteria they intend to report are actually appropriate for the estimation method their chosen software implements.

When writing up SEM results for a thesis, dissertation, or journal submission, Hair et al. (2019b) recommend that authors report, at a minimum: (a) the measurement model results, including indicator loadings, internal consistency reliability, and convergent and discriminant validity statistics, with reference to the threshold values summarized in this article; (b) the structural model results, including path coefficients, their significance levels, and, for CB-SEM, the full set of global fit indices, or, for PLS-SEM, R^2, f^2, and Q^2 values, together with PLSpredict results where prediction is a study objective, and, for models with mediation or moderation, the bootstrapped indirect or interaction effects with their confidence intervals; and (c) the specific software, version, and estimation algorithm used, so that the analysis can be reproduced. Following a consistent, criterion-based reporting structure of this kind not only strengthens the methodological rigor of a researcher's own research but also makes it considerably easier for examiners, reviewers, and readers to evaluate the credibility of the reported model (Hair et al., 2019b; Kline, 2015).

9. Conclusion

Structural equation modeling offers social science researchers a powerful framework for testing complex, theory-driven relationships among latent constructs while explicitly modeling measurement error. As this article has shown, the two dominant SEM traditions, covariance-based SEM and partial least squares SEM, rest on different statistical logics and are therefore evaluated using different, non-interchangeable sets of criteria. CB-SEM, as implemented in AMOS, relies on a wide battery of global goodness-of-fit indices, including CMIN/DF, GFI/AGFI, the baseline-comparison indices (NFI, RFI, IFI, TLI, CFI), parsimony-adjusted indices, RMSEA, information criteria, and Hoelter's critical N, alongside normality diagnostics, to confirm that a theoretically specified covariance structure adequately reproduces the observed data. PLS-SEM, by contrast, is evaluated through a systematic assessment of measurement model reliability and validity, namely loadings, composite reliability, AVE, and HTMT for reflective constructs, and redundancy analysis, indicator collinearity, and indicator weights for formative constructs, complemented by SmartPLS's own model fit report (SRMR, d_ULS, d_G, Chi-square, NFI, and RMS_theta) and structural model criteria emphasizing

explanatory and predictive power, namely R^2, f^2, Q^2, and, for genuine out-of-sample prediction, PLSpredict. For social science researchers and students, understanding these distinctions and applying the correct, well-referenced threshold values summarized in this article, is essential not only for producing methodologically sound theses and publications, but also for critically evaluating SEM-based research encountered in the wider literature. Both approaches can accommodate mediation and moderation, provided that indirect effects are tested with bootstrap confidence intervals and interaction effects are specified with methods appropriate to the estimator. Finally, whichever approach is chosen, the directional paths of a structural equation model represent theoretical assumptions rather than demonstrated causal effects, and the conclusions drawn from them should be no stronger than the research design allows.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

Aguinis, H., Beaty, J. C., Boik, R. J., & Pierce, C. A. (2005). Effect Size and Power in Assessing Moderating Effects of Categorical Variables Using Multiple Regression: A 30-Year Review. *Journal of Applied Psychology*, *90*, 94-107. https://doi.org/10.1037/0021-9010.90.1.94

Akaike, H. (1987). Factor Analysis and AIC. *Psychometrika*, *52*, 317-332. https://doi.org/10.1007/bf02294359

Antonakis, J., Bendahan, S., Jacquart, P., & Lalive, R. (2010). On Making Causal Claims: A Review and Recommendations. *The Leadership Quarterly*, *21*, 1086-1120. https://doi.org/10.1016/j.leaqua.2010.10.010

Bagozzi, R. P., & Yi, Y. (1988). On the Evaluation of Structural Equation Models. *Journal of the Academy of Marketing Science*, *16*, 74-94. https://doi.org/10.1007/bf02723327

Baron, R. M., & Kenny, D. A. (1986). The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations. *Journal of Personality and Social Psychology*, *51*, 1173-1182. https://doi.org/10.1037/0022-3514.51.6.1173

Becker, J. M., Ringle, C. M., & Sarstedt, M. (2018). Estimating Moderating Effects in PLS-SEM and PLSc-SEM: Interaction Term Generationdata Treatment. *Journal of Applied Structural Equation Modeling*, *2*, 1-21. https://doi.org/10.47263/jasem.2(2)01

Bentler, P. M. (1990). Comparative Fit Indexes in Structural Models. *Psychological Bulletin*, *107*, 238-246. https://doi.org/10.1037/0033-2909.107.2.238

Bentler, P. M., & Bonett, D. G. (1980). Significance Tests and Goodness of Fit in the Analysis of Covariance Structures. *Psychological Bulletin*, *88*, 588-606. https://doi.org/10.1037/0033-2909.88.3.588

Bollen, K. A. (1989). A New Incremental Fit Index for General Structural Equation Models. *Sociological Methods & Research*, *17*, 303-316. https://doi.org/10.1177/0049124189017003004

Bollen, K. A., & Pearl, J. (2013). Eight Myths about Causality and Structural Equation Models. In S. L. Morgan (Ed.), *Handbook of Causal Analysis for Social Research* (pp. 301-328). Springer.

Bozdogan, H. (1987). Model Selection and Akaike's Information Criterion (AIC): The General Theory and Its Analytical Extensions. *Psychometrika*, *52*, 345-370. https://doi.org/10.1007/bf02294361

Browne, M. W., & Cudeck, R. (1989). Single Sample Cross-Validation Indices for Covariance Structures. *Multivariate Behavioral Research*, *24*, 445-455. https://doi.org/10.1207/s15327906mbr2404_4

Browne, M. W., & Cudeck, R. (1993). Alternative Ways of Assessing Model Fit. In K. A. Bollen, & J. S. Long (Eds.), *Testing Structural Equation Models* (pp. 136-162). Sage Publications.

Byrne, B. M. (2016). *Structural Equation Modeling with AMOS: Basic Concepts, Applications, and Programming* (3rd ed.). Routledge.

Cenfetelli, R. T., & Bassellier, G. (2009). Interpretation of Formative Measurement in Information Systems Research. *MIS Quarterly*, *33*, 689-708. https://doi.org/10.2307/20650323

Cheah, J. H., Sarstedt, M., Ringle, C. M., Ramayah, T., & Ting, H. (2018). Convergent Validity Assessment of Formatively Measured Constructs in PLS-SEM: On Using Single-Item versus Multi-Item Measures in Redundancy Analyses. *International Journal of Contemporary Hospitality Management*, *30*, 3192-3210. https://doi.org/10.1108/ijchm-10-2017-0649

Chin, W. W. (1998). The Partial Least Squares Approach to Structural Equation Modeling. In G. A. Marcoulides (Ed.), *Modern Methods for Business Research* (pp. 295-336). Lawrence Erlbaum Associates.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Lawrence Erlbaum Associates.

Diamantopoulos, A., & Siguaw, J. A. (2006). Formative versus Reflective Indicators in Organizational Measure Development: A Comparison and Empirical Illustration. *British Journal of Management*, *17*, 263-282. https://doi.org/10.1111/j.1467-8551.2006.00500.x

Diamantopoulos, A., & Winklhofer, H. M. (2001). Index Construction with Formative Indicators: An Alternative to Scale Development. *Journal of Marketing Research*, *38*, 269-277. https://doi.org/10.1509/jmkr.38.2.269.18845

Dijkstra, T. K., & Henseler, J. (2015). Consistent Partial Least Squares Path Modeling. *MIS Quarterly*, *39*, 297-316. https://doi.org/10.25300/misq/2015/39.2.02

Fornell, C., & Larcker, D. F. (1981). Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. *Journal of Marketing Research*, *18*, 39-50. https://doi.org/10.1177/002224378101800104

Geisser, S. (1974). A Predictive Approach to the Random Effect Model. *Biometrika*, *61*, 101-107. https://doi.org/10.1093/biomet/61.1.101

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019a). *Multivariate Data Analysis* (8th ed.). Cengage.

Hair, J. F., Risher, J. J., Sarstedt, M., & Ringle, C. M. (2019b). When to Use and How to Report the Results of PLS-SEM. *European Business Review*, *31*, 2-24. https://doi.org/10.1108/ebr-11-2018-0203

Hayes, A. F. (2015). An Index and Test of Linear Moderated Mediation. *Multivariate Behavioral Research*, *50*, 1-22. https://doi.org/10.1080/00273171.2014.962683

Henseler, J., & Chin, W. W. (2010). A Comparison of Approaches for the Analysis of Interaction Effects between Latent Variables Using Partial Least Squares Path Modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *17*, 82-109.

https://doi.org/10.1080/10705510903439003

Henseler, J., Dijkstra, T. K., Sarstedt, M., Ringle, C. M., Diamantopoulos, A., Straub, D. W., Ketchen, D. J., Hair, J. F., Hult, G. T. M., & Calantone, R. J. (2014). Common Beliefs and Reality about PLS: Comments on Rönkkö and Evermann (2013). *Organizational Research Methods*, *17*, 182-209. https://doi.org/10.1177/1094428114526928

Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A New Criterion for Assessing Discriminant Validity in Variance-Based Structural Equation Modeling. *Journal of the Academy of Marketing Science*, *43*, 115-135. https://doi.org/10.1007/s11747-014-0403-8

Henseler, J., Ringle, C. M., & Sarstedt, M. (2016). Testing Measurement Invariance of Composites Using Partial Least Squares. *International Marketing Review*, *33*, 405-431. https://doi.org/10.1108/imr-09-2014-0304

Hoelter, J. W. (1983). The Analysis of Covariance Structures: Goodness-of-Fit Indices. *Sociological Methods & Research*, *11*, 325-344. https://doi.org/10.1177/0049124183011003003

Hooper, D., Coughlan, J., & Mullen, M. R. (2008). Structural Equation Modelling: Guidelines for Determining Model Fit. *Electronic Journal of Business Research Methods*, *6*, 53-60.

Hu, L., & Bentler, P. M. (1999). Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria versus New Alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*, 1-55. https://doi.org/10.1080/10705519909540118

Jarvis, C. B., MacKenzie, S. B., & Podsakoff, P. M. (2003). A Critical Review of Construct Indicators and Measurement Model Misspecification in Marketing and Consumer Research. *Journal of Consumer Research*, *30*, 199-218. https://doi.org/10.1086/376806

Jöreskog, K. G. (1969). A General Approach to Confirmatory Maximum Likelihood Factor Analysis. *Psychometrika*, *34*, 183-202. https://doi.org/10.1007/bf02289343

Jöreskog, K. G., & Sörbom, D. (1984). *LISREL VI: Analysis of Linear Structural Relationships by Maximum Likelihood, Instrumental Variables, and Least Squares Methods* (3rd ed.). Scientific Software.

Kenny, D. A., & Judd, C. M. (1984). Estimating the Nonlinear and Interactive Effects of Latent Variables. *Psychological Bulletin*, *96*, 201-210. https://doi.org/10.1037/0033-2909.96.1.201

Klein, A., & Moosbrugger, H. (2000). Maximum Likelihood Estimation of Latent Interaction Effects with the LMS Method. *Psychometrika*, *65*, 457-474. https://doi.org/10.1007/bf02296338

Kline, R. B. (2015). *Principles and Practice of Structural Equation Modeling* (4th ed.). Guilford Press.

MacCallum, R. C., Wegener, D. T., Uchino, B. N., & Fabrigar, L. R. (1993). The Problem of Equivalent Models in Applications of Covariance Structure Analysis. *Psychological Bulletin*, *114*, 185-199. https://doi.org/10.1037/0033-2909.114.1.185

Mardia, K. V. (1970). Measures of Multivariate Skewness and Kurtosis with Applications. *Biometrika*, *57*, 519-530. https://doi.org/10.1093/biomet/57.3.519

Marsh, H. W., Wen, Z., & Hau, K. T. (2004). Structural Equation Models of Latent Interactions: Evaluation of Alternative Estimation Strategies and Indicator Construction. *Psychological Methods*, *9*, 275-300. https://doi.org/10.1037/1082-989x.9.3.275

Maxwell, S. E., & Cole, D. A. (2007). Bias in Cross-Sectional Analyses of Longitudinal Mediation. *Psychological Methods*, *12*, 23-44. https://doi.org/10.1037/1082-989x.12.1.23

Nitzl, C., Roldán, J. L., & Cepeda, G. (2016). Mediation Analysis in Partial Least Squares Path Modeling: Helping Researchers Discuss More Sophisticated Models. *Industrial*

Management & Data Systems, *116*, 1849-1864. https://doi.org/10.1108/imds-07-2015-0302

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric Theory* (3rd ed.). McGraw-Hill.

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and Resampling Strategies for Assessing and Comparing Indirect Effects in Multiple Mediator Models. *Behavior Research Methods*, *40*, 879-891. https://doi.org/10.3758/brm.40.3.879

Sarstedt, M., Ringle, C. M., & Hair, J. F. (2017). Partial Least Squares Structural Equation Modeling. In C. Homburg, M. Klarmann, & A. Vomberg (Eds.), *Handbook of Market Research* (pp. 1-40). Springer. https://doi.org/10.1007/978-3-319-05542-8_15-1

Shmueli, G., Ray, S., Velasquez Estrada, J. M., & Chatla, S. B. (2016). The Elephant in the Room: Predictive Performance of PLS Models. *Journal of Business Research*, *69*, 4552-4564. https://doi.org/10.1016/j.jbusres.2016.03.049

Shmueli, G., Sarstedt, M., Hair, J. F., Cheah, J. H., Ting, H., Vaithilingam, S., & Ringle, C. M. (2019). Predictive Model Assessment in PLS-SEM: Guidelines for Using PLSpredict. *European Journal of Marketing*, *53*, 2322-2347. https://doi.org/10.1108/ejm-02-2019-0189

Steiger, J. H. (1990). Structural Model Evaluation and Modification: An Interval Estimation Approach. *Multivariate Behavioral Research*, *25*, 173-180. https://doi.org/10.1207/s15327906mbr2502_4

Stone, M. (1974). Cross-Validatory Choice and Assessment of Statistical Predictions. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *36*, 111-133. https://doi.org/10.1111/j.2517-6161.1974.tb00994.x

West, S. G., Finch, J. F., & Curran, P. J. (1995). Structural Equation Models with Nonnormal Variables: Problems and Remedies. In R. H. Hoyle (Ed.), *Structural Equation Modeling: Concepts, Issues, and Applications* (pp. 56-75). Sage Publications.

Wold, H. (1982). Soft Modeling: The Basic Design and Some Extensions. In K. G. Jöreskog, & H. Wold (Eds.), *Systems under Indirect Observation: Part 2* (pp. 1-54). North-Holland.

Zhao, X., Lynch Jr, J. G., & Chen, Q. (2010). Reconsidering Baron and Kenny: Myths and Truths about Mediation Analysis. *Journal of Consumer Research*, *37*, 197-206. https://doi.org/10.1086/651257