ANNUAL REVIEWS

Annual Review of Organizational Psychology and Organizational Behavior

Structural Equation Modeling in Organizational Research: The State of Our Science and Some Proposals for Its Future

Michael J. Zyphur,[1] Cavan V. Bonner,[2] and Louis Tay[2]

[1]UQ Business School, University of Queensland, St Lucia, Queensland, Australia; email: m.zyphur@uq.edu.au

[2]Department of Psychological Sciences, Purdue University, West Lafayette, Indiana, USA

Annu. Rev. Organ. Psychol. Organ. Behav. 2023. 10:495–517

First published as a Review in Advance on November 14, 2022

The *Annual Review of Organizational Psychology and Organizational Behavior* is online at orgpsych.annualreviews.org

https://doi.org/10.1146/annurev-orgpsych-041621-031401

Keywords

structural equation model, structural equation modeling, SEM, model fit, latent interactions, latent variable models

Abstract

The use of structural equation modeling (SEM) has grown substantially over the past 40 years within organizational research and beyond. There have been many different developments in SEM that make it increasingly useful for a variety of data types, research designs, research questions, and research contexts in the organizational sciences. To give researchers a better understanding of how and why SEM is used, our article (*a*) presents a review of SEM applications within organizational research; (*b*) discusses SEM best practices; and (*c*) explores advanced SEM applications, including instrumental variable methods, latent variable interactions and nonlinear measurement models, multilevel SEM, cross-lagged panel models and dynamic structural equation models, and meta-analytic SEM. We conclude by discussing concerns and debates that are both methodological (i.e., cross-validation and regularization) and theoretical (i.e., understanding causal evidence) as they relate to SEM and its application in organizational research and beyond.

STRUCTURAL EQUATION MODELING IN ORGANIZATIONAL RESEARCH: STATE OF THE SCIENCE

The use of structural equation modeling (SEM) has grown tremendously within the social sciences over the past 40 years (Tarka 2018). This can be seen in multiple books, reviews, software, and tutorials on SEM (e.g., Bollen & Long 1993, Hoyle & Smith 1994, Kline 2004, Weston & Gore 2016), with even a methodological journal, *Structural Equation Modeling: A Multidisciplinary Journal*, devoted to this method since 1994. As the complexity of SEM and its applications increase, there is a need to better understand it within the context of a field that has unique demands in terms of research designs, data, contexts, and questions. In this regard, we seek to critically evaluate how SEM has been applied within the field of organizational research.

There are three main goals of our article. Foremost, we review classic SEM applications within organizational research to understand the trends in SEM use while also examining reporting practices. This allows us to provide practical recommendations for researchers using SEM. Second, we review increasingly popular advanced SEM applications such as multilevel and latent interaction modeling, and we provide resources for applying them. This goes beyond the landmark review by MacCallum & Austin (2000) in the general field of psychology, allowing us to address concerns of greater transparency and replicability through better SEM reporting while also addressing significant advances in SEM over the past two decades. Third, we conclude by raising important issues related to how structural equation models are empirically evaluated in relation to cross-validation, regularization, and machine learning, but also by focusing on the concepts of evidence and causality. Our aim is to make points that are somewhat contentious and, in turn, to spur important new debates about structural equation models and how they can be evaluated and their results applied in real-world contexts. Taken together, these three contributions will allow researchers, especially doctoral students, to better understand the nature and historical applications of SEM, as well as the ways that advanced SEM techniques can be usefully applied, while simultaneously being offered conceptual tools for critically evaluating structural equation models and their applications.

Overview of Structural Equation Modeling

As excellent resources already exist on introductory SEM topics (e.g., Kline 2004), we provide only a brief primer here, emphasizing the basic concepts and advantages of applying SEM rather than explaining the statistical estimation theory underlying it. SEM refers to a set of statistical techniques that enable modeling relationships between multiple independent variables (IVs) and dependent variables (DVs)—these are also often called exogenous and endogenous variables or predictors and criteria/outcomes, respectively. One key advantage of SEM is that these IVs and DVs can be modeled as observed variables that are realized in a dataset or latent variables that must be estimated based on observed data (Bollen 2002). Standard SEM is typically thought of as a combination of two approaches: (*a*) a structural model that specifies how the substantive IVs causally relate to the DVs (i.e., path analysis); and (*b*) a measurement model that specifies how the observed indicator variables (e.g., self-report scale items) are related to the latent variables (i.e., factor analysis), wherein the latent variables can be understood as causing the responses on the observed variables. While typically examined in combination, each type of model can also be analyzed and its fit assessed separately in a structural equation model (see McDonald 2010, O'Boyle & Williams 2011).

To illustrate this, we graphically depict a structural equation model in **Figure 1**, which shows three latent variables, X, M, and Y, forming an indirect effect (i.e., mediational path). X is an exogenous latent IV, whereas M and Y are endogenous latent DVs. The structural model is shown

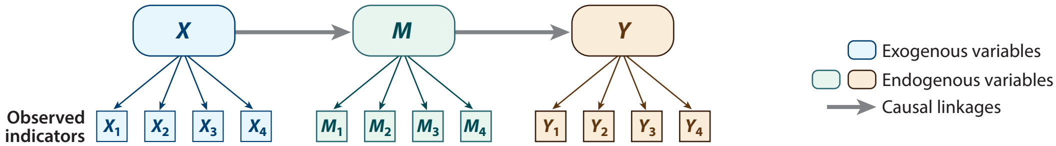

Figure 1

Illustrative structural equation model: mediation analysis. *X* is an exogeneous variable; *M* and *Y* are both endogenous variables. As part of the measurement model, thin single-headed arrows represent measurement paths connecting the latent variables to their respective observed indicators. As part of the structural model, thick single-headed arrows represent the causal linkages connecting the latent variables.

in the thick single-headed arrows representing causal connections among the latent variables; the measurement model is shown in the thin single-headed arrows connecting the latent variables to the observed variables. In both cases, the models and estimation can be understood as a type of regression analysis with predictors and outcomes—as we often say when teaching SEM and other modeling topics, "It's all just regression."

There are many conceptual reasons why SEM has gained popularity over the years. First, SEM is not merely a methodological tool for data analysis; SEM also serves as a way to theoretically conceptualize and empirically test complex propositions or hypotheses that are substantive in nature (see Zyphur 2009). These include mediation models (e.g., Quigley et al. 2020, Solomon et al. 2021, Wanberg et al. 2020) and examining distinct indirect-effects pathways for different IV-DV relationships (e.g., Livne-Ofer et al. 2019, Loi et al. 2020, Zhang et al. 2020). Second, SEM provides a formal way to test different alternative models that offer different conceptual pictures of the world, along with clear statistical criteria that allow evaluating whether one model fits better than another—thereby allowing researchers to adjudicate between competing theories and hypotheses. For instance, Fulmer & Ostroff (2017) compared models that specified a trickle-up versus trickle-down causal structure for the relationship between trust in direct leaders, top leaders, and job performance. Third, SEM is a tool that is flexible enough to be applied not only to nonexperimental observational data but also to experimental research designs such as randomized control trials (RCTs) that can be mapped somewhat directly to counterfactual theories of causal inference (for overviews, see Muthén & Asparouhov 2015, Muthén et al. 2017, Pearl 2010). Fourth, through measurement models, SEM provides a way to estimate relationships between constructs that account for (un)reliability in the observed indicator variables. This enables organizational researchers to assess theoretical relationships that are often of interest rather than observed associations that are biased by measurement error variance (Viswesvaran & Ones 1995).

More generally, specifying measurement models within SEM forces researchers to think carefully about the type of latent construct they are assessing (Williams et al. 2009). In particular, the traditional and typical approach to SEM involves constructs that are modeled as reflective (arrows traveling from the latent construct to the indicators), to be contrasted with models of formative constructs (arrows traveling from indicators to latent constructs) (Edwards & Bagozzi 2000). We primarily focus on the former in our review but will compare both in a subsequent section.

Standard Structural Equation Modeling in Organizational Research

To better understand recent trends in SEM within organizational research, we conducted a review of standard SEM applications in our top-ranked journals. Specifically, we identified articles with the keywords "structural equation model*" or "SEM" over the past 10 years in the *Academy of Management Journal* and *Journal of Applied Psychology*. This search returned a total of 668 articles. For our purposes, we were only interested in standard SEM applications that included both a

Table 1 Summary of standard structural equation modeling (SEM) functions and issues to consider

Functions	Application(s)	Issues to consider
Modeling and testing complex models	Model relationships among different types of constructs and variables. Model multiple indirect pathways (e.g., mediational models) and multiple predictors and outcomes.	Highly complex models are less replicable. Estimating multiple indirect effects can be compromised when they are correlated. Should seek to validate complex SEM in holdout samples or additional samples
Model comparisons	Test and compare multiple competing theoretical models. Test for a more parsimonious model compared to the one posited.	Be mindful of multiple statistically equivalent models for a given theory. Researchers should seek to rule out alternative plausible theoretical models using SEM.
Modeling experimental designs	Model experimental effects controlling for measurement error. Model causal processes in experiments. Test for measurement and structural equivalence between design conditions.	Systematic error variance will not be addressable with SEM but will require consideration of measurement and experimental procedures.
Modeling theoretical relationships	Estimate theoretical relationships of constructs controlling for measurement error. Determine if theoretical measurement properties hold across different groups (e.g., gender, race, nationality) and over time. Model reflective and/or formative constructs.	Models with many latent variables can be challenging to estimate together and may require a piecemeal or two-step approach. Traditional conceptions of reliability may not apply to formative constructs.

structural and measurement model. Therefore, we excluded applications where there was only a measurement model [i.e., confirmatory factor analysis (CFA)] and applications where there was only a structural model (e.g., path analysis among observed variables). In addition, because we were interested in reviewing more advanced applications (which we do in subsequent, separate sections), we excluded more sophisticated approaches from our initial review, including latent interactions, multilevel SEM (MSEM), and Bayesian estimation.

After applying these exclusion criteria and removing duplicates, we were left with 97 articles. Some articles included multiple studies that tested separate structural equation models. Because we are focused on general reporting practices, we only coded the first study in multistudy papers. In the case of studies that reported multiple equivalent structural equation models [e.g., five simultaneous models for each of the Big 5 traits (Connelly et al. 2022)], we only reported the first model. In the case of studies that reported multiple model iterations that were successively (re)specified after estimation, we focused on what the authors deemed to be their final model. Because we are interested in general reporting practices, we present all data at the article level, not the study or model level.

In what follows, we critically evaluate some of the SEM applications and reporting practices based on the advantages, or the key functions, that SEM is supposed to offer researchers and the research community. We summarize this in **Table 1**. The set of final articles included in our review can be found on the Center for Open Science's OSF platform (**https://osf.io/kqe2p/**).

Model complexity. Theories of organizations and the people who inhabit them have become increasingly complex over time by incorporating features such as temporal dynamics (Ancona et al. 2001), digital technologies, and globalization (Baum & Haveman 2020). The result has been an increase in the extent to which our theories are able to integrate multiple perspectives and data types. To properly address such features of theory requires corresponding methodological tools that can empirically assess theoretical propositions and the operational hypotheses that follow from them. For this purpose, SEM's flexibility has made it an increasingly popular tool for assessing and testing complex theoretical models. These include multiple indirect-effects pathways as well as, potentially, feedback loops over time. The key to these model features is that SEM allows going beyond traditional regression analysis wherein only a single DV is predicted by some set of IVs—more formally, traditional regression is a univariate technique, whereas SEM is multivariate. For example, Wo et al. (2015) used SEM to capture multiple mediating mechanisms that cause trickle-down effects, with supervisor justice perceptions influencing their subordinates' justice perceptions.

However, there are several issues that researchers need to be mindful of when testing complex models. Foremost, recent research has recognized that models with multiple variables have substantially lower probabilities of being correct. Assuming an 80% probability of each correlation being properly theorized (perhaps in a rudimentary yes/no fashion), the joint probability of a six-variable model being accurate is approximately 3.5% (Saylors & Trafimow 2020). This problem of larger models with more complexity being less likely to be accurate is exacerbated when models include causal chains with multiple intervening mediators. Statistically significant results in such models may be due merely to chance and have low replicability. Therefore, when specifying complex models, we encourage researchers to be mindful of the potential base rates of being wrong (Saylors & Trafimow 2020). As a method of addressing this and related issues of model specification and generalization, we point out that validation approaches could be used, wherein a proposed model is validated in a holdout sample (estimating a structural equation model with a training dataset and then evaluating its fit with a test dataset) or a different sample altogether. There were very few instances of traditional cross-validation with a holdout sample among the articles reviewed, and no applications of k-fold cross-validation, which has the added advantage of validating the model using the entire dataset by sequentially using each data subset in the training/test set. For example, Nye et al. (2014) randomly split their data into calibration and validation samples to test the robustness of their measurement model, in addition to testing their structural model of sexual harassment with a second, separate sample.

With complex models, researchers often seek to test multiple indirect effects. For example, Priesemuth et al. (2014) applied SEM to test a model where abusive supervision climate indirectly influenced multiple DVs through two distinct mediating mechanisms. Statistically, to the extent that these indirect effects are correlated, it can affect the accuracy of the results. One possible solution is to orthogonalize the mediators to reduce this problem (see related thought in Zyphur et al. 2020a).

Model comparisons. Another advantage of SEM is the ability for researchers to test for competing explanations by testing multiple alternative models. From a Popperian perspective, the scientific process requires researchers to engage in tests that may falsify one or another model by pitting models against each other. From a pragmatic perspective, testing equally viable alternative models can reduce confirmation bias. Despite this, the process of choosing specific alternative models to compare against a researcher's preferred model is often challenging. While some papers do provide tests for alternative models (e.g., Ferris et al. 2016, Fulmer & Ostroff 2017, Kunze et al. 2015), a substantial number of papers do not make SEM model comparisons.

In an excellent treatment of this topic, Vandenberg & Grelle (2009) note that it is critical to determine the types of SEM models that one seeks to test in the first place. The first type is equivalent models, where alternate models typically produce an identical overall goodness-of-fit to the data. For instance, a proposed model of $X \rightarrow M \rightarrow Y$ would have an equivalent model (in terms of fit) of $Y \rightarrow M \rightarrow X$. It has been estimated that there are a substantial number of equivalent models based on published applications, with half of the applications having 16 or more equivalent models (MacCallum et al. 1993). To engage in strong tests of theory through model comparisons, it is vital for researchers to identify plausible equivalent alternatives and develop research designs to rule these out, including through the use of experimental methods and longitudinal data collection, rather than relying on the "status of an a priori model" (MacCallum et al. 1993, p. 197). By seeking to test different equally plausible model-based theories, we are not merely utilizing SEM in an ad hoc exploratory fashion but also seeking to critically evaluate models in a more thoughtful confirmatory manner—the process of ruling out alternatives is part of what differentiates confirmatory from exploratory research.

Second, most structural equation models involve the possibility of nested model comparisons, wherein alternative models have some subset of the parameters that are estimated in the posited model (Vandenberg & Grelle 2009), which allows direct statistical tests to evaluate whether the nested model fits more poorly than the posited model (Bentler & Satorra 2010). For example, one may posit that outcome Y is being predicted by three IVs, but a nested model may only involve two of the three predictors' paths being freely estimated. Simpler, counterfactual models can test the assumptions underlying the initial model. For example, Foulk et al. (2016) initially tested a SEM where the rude behavior of a source participant in a negotiation task had a contagious effect on the rudeness of a carrier participant, such that a third participant's perception of the carrier's rudeness during a second negation task was indirectly affected by the source's contagious rudeness. Foulk et al. (2016) found that a nested model where the path from the source to the mediating carrier participant was eliminated was significantly worse fitting than the full three-participant mediation model, which supported the importance of the initial source's rudeness.

Because nested models are more parsimonious by definition, they will often reflect simpler and potentially more elegant theories. This makes alternative nested models potentially more desirable because they involve a lower probability of being wrong by reducing model complexity and, as per Occam's Razor, may be more aesthetically pleasing to the scientific community while at the same time reduce the chances of overfitting a structural equation model to the data.

Finally, there are non-nested models where the variables in the analysis may be overlapping but nonequivalent (Vandenberg & Grelle 2009). For example, a researcher may posit that job satisfaction mediates the relationship between personality and job performance; however, an alternative non-nested model might include organizational commitment as the mediator. In this case, researchers may use model fit indices such as the Akaike information criterion (AIC) and the Bayesian information criterion (BIC) to compare alternative models [where the same data and variables are used (see Vrieze 2012 for more information on when to use AIC or BIC)]. In this regard, researchers need to consider alternative explanations and variables where possible. For example, McCarthy et al. (2016) used these indices to compare several alternative non-nested structural models of the relationship between workplace anxiety and job performance to their hypothesized model.

Experimental designs. While the use of SEM has grown over time, the applications have typically been based on observational data that do not involve experimental interventions, including cross-sectional and longitudinal data. Among the articles that we reviewed, 62% applied SEM to cross-sectional data, 38% to longitudinal data, and just 4% were applied to experimental designs. However, SEM can easily be used to analyze data from experiments (MacCallum & Austin 2000).

For example, Hershcovis & Bhatnagar (2017) used SEM to test a mediation model with data from a traditional vignette experiment to investigate how customers respond to employee mistreatment. Unlike a standard *t*-test or analysis of variance approach to infer group mean differences from observed scores, SEM can account for measurement error variance and thus help to address unreliability in observed variables, which can otherwise bias estimates of treatment effects (although note that systematic errors, such as a construct deficient measure, cannot be addressed by the use of SEM without incorporating additional information into the model, such as with instrumental variables as we describe further below). Also, SEM can be used to model mediating causal processes in experiments directly (Russell et al. 1998), and it allows accounting for correlated measurement errors over time in pre- and post-test experimental designs that use the same measures (Ployhart & Oswald 2004). Also important is that SEM allows measurement invariance tests to evaluate whether observed variables are measuring the same constructs in different groups or over time; when this is not found, researchers can still specify partial measurement invariance and continue testing their hypothesized effects (Byrne et al. 1989). In general, given advantages such as these, we encourage researchers to use SEM in the analysis of experimental data, which has become increasingly popular recently including for causal mediation modeling (see Muthén & Asparouhov 2015).

Theoretical relationships. A key purpose of SEM is estimating the relationships among theoretical constructs directly as latent variables, rather than among observed indicator measures [e.g., operational relationship (Binning & Barrett 1989)]. To this end, SEM is often described as a method that accounts for measurement error variance. Nevertheless, there are often situations wherein researchers rely on a two-stage piecemeal approach due to difficulties estimating all of the latent variables and relationships that a given model may require, especially in cases when there are many observed or latent variables and/or small samples.

In such cases, the estimation stages are typically (*a*) show that the measurement model holds for the measures of interest and (*b*) use observed scores for constructs within a path analytic framework to assess the theoretical relationships. For example, in a recent study assessing the effects of work-contingent self-esteem, the authors first fit an 11-factor measurement model to all the scales in order to validate their factor structure and then, second, proceeded to use a mediational path model using scale means as observed variables (Kuykendall et al. 2019). Unfortunately, this approach has historically not taken full advantage of SEM in accounting for measurement error and correctly reflecting uncertainty in latent variable standings, which is a problem that becomes more pronounced as scales become shorter and less reliable. We propose that when it is challenging to conduct SEM due to the number of latent variables, researchers might use factor scores within a structural model but should be very mindful of the potential shortcomings of this method (see Hayes & Usami 2020, Skrondal & Laake 2001).

There are two possibilities for addressing the problems with factor scores. One would be to use an estimation strategy that attempts to overcome the deficiencies of factor score regressions, such as the Structural-After-Measurement estimation method implemented in the lavaan package for R (see Rosseel & Loh 2021). A second possibility is to use plausible values (as Bayesian posterior estimates) rather than factor scores or observed mean scores, because plausible values take into account uncertainty in the standings on latent variables (see Zyphur et al. 2019). This approach is implemented both in Mplus and for lavaan in R, with the latter made possible through the semTools package using the plausibleValues( ) function (Asparouhov & Muthén 2010, Jorgensen et al. 2022).

It is helpful to note that the two-stage piecemeal process is related to the classic Anderson & Gerbing (1988) two-step modeling approach, where it is recommended that a measurement model is specified in the first step (i.e., a multiple factor model with fully saturated covariances among the latent variables) before specifying and evaluating the structural paths in the second step. The goal of using these two steps is to reduce interpretational confounding, which occurs when the measurement model (parameters and fit) depends on the specification of structural paths. Estimating a measurement model first and ensuring no misspecification in a first step facilitates meaningful interpretations of subsequent structural model estimates in a second step (i.e., regressions among latent variables).

Importantly, in the second step, Anderson & Gerbing (1988) propose evaluating the structural paths in a full SEM that also includes the measurement model. This is distinct from, although related to, the aforementioned piecemeal approach of estimating the measurement model first and then the structural model later using scale means, factor scores, or plausible values. We recommend the Anderson & Gerbing (1988) two-step modeling approach when possible rather than a piecemeal approach to test theoretical relationships. In either case, like Anderson & Gerbing (1988) we also advocate evaluating the posited theoretical model by comparing it to one or more structural models that are more constrained (i.e., fewer structural paths to test for a more parsimonious model) and less constrained (i.e., more structural paths to test if the posited theoretical model is potentially too parsimonious).

An interesting advantage of SEM is that it not only allows assessing theoretical causal relationships, but it also allows determining whether the theoretical measurement properties of scales hold across multiple populations (e.g., gender, race, nationality). This type of classic analysis of measurement invariance is actually a method of testing the theoretical relationships among scale items in the form of the measurement relationships among latent variables and their observed indicators. Such tests of measurement equivalence should be done when researchers are assessing theoretical constructs with measurement models across multiple populations and/or over time (Tay et al. 2015, Vandenberg & Lance 2000). To this end, recent advances allow going beyond traditional multigroup invariance testing and similar tests over time by allowing for approximate measurement invariance tests, the results of which can allow some degree of noninvariance to be observed while still allowing for the substantive tests that researchers may be seeking to conduct after an initial observation of similar measurement characteristics across groups and/or over time (Davidov et al. 2018, Muthén & Asparouhov 2017).

More generally, the specification of measurement models in SEM motivates researchers to consider the nature of the latent constructs they are modeling, such as whether they are formative or reflective in the context of a given set of observed indicators, as depicted in **Figure 2**. Reflective measurement is the most typical approach, as previously mentioned. Williams et al. (2009) note that a reflective construct might be mapped to a realist epistemology, wherein measures are

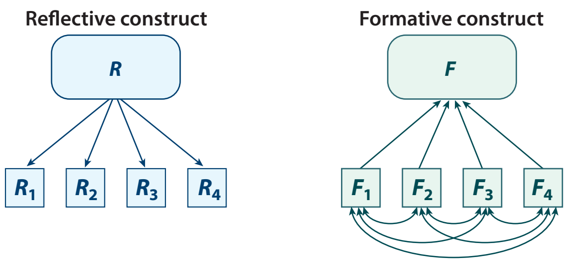

Figure 2

Reflective versus formative construct, where the difference in the direction of arrows indicates the different types of constructs (arrows from a latent variable to observed indicators is reflective; arrows to a latent variable from observed indicators is formative).

imperfectly assessing a “real” entity that exists beyond the measurement tool—thus, a “construct” nonironically actually exists in the world. For example, a self-report measure of job performance is assumed to be a fallible assessment of “true” job performance, which is behavioral and actually done as a set of actions in the world.

On the other hand, a formative construct might be mapped to a constructivist epistemology, wherein the construct is formed as a purely conceptual feature of the world by appealing to the indicators that are used to produce images of it. A classic example is socioeconomic status, which is comprised of a set of only partially related indicators (e.g., income, education, family size); also consider, however, that supervisor-rated job performance in different domains may be regarded as forming an overall image of job performance, and the question of the truth of the image is irrelevant from a realist perspective because this typically is not adopted epistemologically. More generally, researchers have noted that job performance may be an exemplar of a formative construct in the organizational sciences because (*a*) changes in different domains (e.g., task performance, organizational citizenship behavior, and counterproductive work behavior) of job performance can change overall job performance, (*b*) different aspects of job performance may have different antecedents, and (*c*) different domains of job performance are unique and may not correlate with other domains (MacKenzie et al. 2005).

As the reader can intuit, there are philosophical issues in terms of how constructs can be understood that are beyond the scope of our article (see realist perspectives offered in Borsboom et al. 2003, Edwards & Bagozzi 2000; see alternatives in Zyphur & Pierides 2017, 2020a,b; Zyphur et al. 2015). However, in terms of practical implications, we point out that unlike reflective constructs, indicators of formative constructs are essential and not typically exchangeable, as each indicator uniquely forms the meaning of the construct (Petter et al. 2007). Furthermore, the traditional conceptions of reliability, such as internal consistency reliability, do not apply to formative constructs (Edwards 2010). Finally, there are additional requirements of specifying formative constructs in order for the measurement model to be identified and thus the parameters estimable (Petter et al. 2007, Williams et al. 2009), and there are ways of interpreting the effects of formative constructs that allow an indirect link back to the original indicators themselves—the formative construct is akin to a block of predictor variables (Edwards & Bagozzi 2000).

Structural Equation Modeling Reporting in Organizational Research

With increased concern about transparency and replicability, we offer some specific recommendations for SEM reporting based on trends in organizational research. This augments general SEM reporting guidelines and recommendations (see McDonald & Ho 2002, Raykov et al. 1991, Schreiber et al. 2010). To summarize, we emphasize (*a*) completeness in reporting how a structural equation model was specified and analyzed in terms of model presentation (e.g., using full graphical presentations) and model specification process (e.g., type of data analyzed, missing data and treatment, normality assumptions, and estimator type) and (*b*) the need for a thorough evaluation of the SEM, including the use of multiple appropriate fit indices to evaluate the model [e.g., comparative fit index (CFI), Tucker-Lewis index (TLI), root mean square error of approximation (RMSEA), standardized root mean square residual (SRMR), AIC, and BIC], reporting detailed parameter estimates, possible model modifications, and comparisons with potential alternative models. In the following, we touch on some of these aspects.

Structural equation modeling specification and analysis. We found that 81% of the papers we reviewed provided full graphical presentations of their model. However, within this subset, there were few papers that provided full and clear information such that the number of parameters could be easily determined from the graphical presentation. While an interested researcher might be able to infer this from the degrees of freedom and sample size used during model estimation, it is important for researchers to report such basic information explicitly in all cases.

All papers, with the exception of one, reported basic descriptive statistics for the model variables (i.e., means, standard deviation, and correlation matrix). However, a minority of papers (6%) included descriptive information about whether multivariate normality assumptions were met (i.e., Mardia's coefficients of skewness and kurtosis). We conjecture that this is because few authors checked. Similarly, only 34% of papers discussed how missing data were addressed and treated. This is concerning because it is likely that most data collections involve some type of missingness. In general, we strongly recommend that such basic information should be carefully examined and offered, and any methods for treating missingness, including full-information estimation or multiple imputation, should be mentioned when they are used. For example, we found that only 40% of papers explicitly mentioned the estimator that was used, which is a very small number. This should be reported for the sake of replicability and so that readers (and reviewers) can critically evaluate an estimator's assumptions.

Structural equation modeling evaluation. In terms of the main SEM fit indices that were reported, the majority of articles reported chi-square (78%), degrees of freedom (78%), CFI (80%), and RMSEA (73%). Less than half reported SRMR (46%), TLI (34%), AIC (13%), and incremental fit index (10%), and less than 10% reported BIC (9%), goodness-of-fit index (7%), normed fit index (NFI) (7%), adjusted goodness-of-fit index (4%), root mean square residual (4%), and deviance information criterion (3%). Despite the moderate to high percentages for some fit indices such as the chi-square, CFI, RMSEA, SRMR, and TLI, our field needs to improve in the reporting of these key fit indices, given the importance of model fit as a measure of model quality. Authors should provide such basic fit information, and if they do not then reviewers and editors should insist on it.

In our review, we found that many papers refer to fit indices cutoffs that were used by previous papers rather than the source of the recommended cutoffs themselves. Less than half (36%) of the papers included references to specific cutoffs. Hu & Bentler (1999) was the most popular cutoff citation (15%), followed by Browne & Cudeck (1993) (10%), and Bentler & Bonett (1980) (4%). We encourage researchers to make specific references to the cutoffs that they are using. Hu & Bentler (1999) recommended cutoffs for CFI ($\geq$ 0.95), TLI ($\geq$ 0.95), SRMR ($\leq$ 0.08), and RMSEA ($\leq$ 0.06) based on a simulation study. They also recommended cutoffs for indices that are rarely presented for contemporary models, such as the relative noncentrality index and McDonald's centrality index. Browne & Cudeck (1993) recommended that an RMSEA value of less than 0.05 indicates "close fit," less than 0.08 indicates "reasonable fit," and greater than 0.1 indicates that "we would not want to employ [the] model" (p. 239). Bentler & Bonett (1980) recommended a more liberal cutoff of $\geq$0.90 for both the NFI and non-normed fit index (also known as the TLI). For an emerging approach, see recent work on dynamic fit indices and their cutoffs, which are tailored to a specific model and dataset and which we recommend the reader consider. (At the time of this writing, a web-based implementation could be found at **https://dynamicfit.app/connect/**; also see McNeish & Wolf 2021.)

We did not find papers that provided more detailed parameter estimates of SEM analyses. While we believe that this is because many articles have page limits, and this may be deemed unnecessarily detailed, we advocate for putting these materials in an online supplement where possible. With regard to whether models were modified from the initial proposed model or if comparisons of alternate models were undertaken, we are interested not so much in the extent this was done in the literature but in whether this was reported. This is because not all modifications or model comparisons are needed. For example, one may seek to test a specific proposition

Table 2 Highlights of advanced structural equation modeling (SEM) applications

Type of SEM	Application(s)	Strength(s)
Instrumental variable methods	Modeling causality in a nonexperimental research design	Addresses potential confounds in causal inference
Latent interactions and nonlinear measurement models	Estimating interactions between latent constructs Estimating nonlinear interactions between latent constructs and observed variables	Advanced estimation techniques that can be applied to high-dimensional data that overcome potential precision and convergence issues
Multilevel SEM	Applying SEM to nested data	Goes beyond traditional multilevel regression models to address nested structure in predictors and related measurement error Estimates level-specific structural and measurement models
Cross-lagged panel models and dynamic structural equation models	Applying SEM to longitudinal panel data and intensive longitudinal data	Making causal inferences while controlling for stable confounding factors over time (e.g., individual differences)
Meta-analytic SEM	Applying SEM to meta-analyses	Modeling measurement and structural parameters to obtain more accurate estimates of effects and effect sizes

in a hypothetico-deductive manner with an SEM analysis. Regardless, our field should seek to report if any SEM modifications or comparison (or none) was made. This can help improve transparency in SEM analyses and the evaluation of these models. More generally, careful and thorough reporting has also been advocated, given the degrees of freedom researchers have in specifying and modifying SEM (see Cortina et al. 2017).

Advanced Structural Equation Modeling Applications

A variety of profound advancements in modeling and estimation methods for SEM have occurred in the past two decades. Here we cover what we believe are the most important contributions that have been made by methodologists who have worked to provide new tools and insights that can be put to use by the average researcher for their substantive work. This includes instrumental variable methods; latent interactions and nonlinear measurement models; MSEM; longitudinal approaches, including cross-lagged panel models (CLPMs) (i.e., panel vector autoregressive models) and the new dynamic structural equation model; and meta-analytic SEM (MA-SEM) methods that are becoming increasingly popular. This is summarized in **Table 2** and we address each in turn.

Instrumental variable methods. The use of SEM to estimate structural relationships among substantive variables does not mean that researchers can safely infer that these estimated relationships are causal (Bullock et al. 1994). Although path analysis has at times erroneously been referred to as causal modeling (Denis & Legerski 2006), the fact is that SEM is only a type of regression analysis, and therefore, typical concerns about causality are not overcome without addressing significant issues related to research design and model specification (Diener et al. 2022).

As one way to address these concerns, recent work has focused on the concept of an instrumental variable for addressing potential confounds in nonexperimental research designs (see Antonakis et al. 2010, Grace 2021, Maydeu-Olivares et al. 2020). The idea behind instrumental variable methods is that researchers can eliminate potential confounders affecting predictor variables by using a type of randomization device, allowing them to decompose a predictor variable into two components: one that is truly exogenous in order to estimate its causal effect on an outcome variable, and a second that is endogenous and therefore not used to estimate a causal effect on an outcome variable.

Historically, two-stage least squares methods have been used for instrumental variable analysis (Step 1: decompose the predictor variable; Step 2: estimate the effect of the exogenous component on the outcome variable). However, SEM allows the required decomposition and estimation of the causal effect of the exogenous component in a single step (Antonakis et al. 2010; see early examples in Frone et al. 1992, 1994). The approach is surprisingly simple: An instrumental variable is allowed to directly affect only the predictor, and in turn (*a*) the predictor's residual becomes the model-estimate endogenous component and is allowed to freely covary with the outcome variable to capture the potential noncausal association, and (*b*) the predictor's direct effect on the outcome variable is then able to reflect a causal association. This simultaneous estimation of an effect and covariance among the predictor and outcome variable would not usually be identified, but this identification is made possible by the instrumental variable not having a direct effect on the outcome variable. By assessing model fit (before including the residual covariance in the model specification), a researcher is able to evaluate some of the instrumental variable model assumptions, and the residual covariance among the predictor and the outcome can be used to infer potential endogeneity driving any noncausal association. The net result is a compelling method for inferring causation in the kind of nonexperimental data that are typically analyzed using SEM, especially given that in SEM, latent variables can be used for the instrumental variable analysis in order to account for measurement error—valid instruments, whether observed or latent (they can be either), should not be affected by measurement error, which is one reason why latent instrumental variables in SEM may be so compelling.

In the coming decades, we anticipate instrumental variable-based SEM to become increasingly popular. In our view, gone are the days when purely observational data could be uncritically subjected to structural model specifications and causality inferred. Authors, reviewers, and the general public are coming to expect more from the social sciences in terms of the veracity of their causal claims and their reliability. Instrumental variables offer one potential approach for addressing issues of causality, and we advocate for their use in the organizational sciences to complement the many other approaches that can also be useful.

Latent interactions and nonlinear measurement models. Another significant advancement over the past two decades is in the area of nonlinear latent variable models, including statistical interactions among latent variables (i.e., latent moderation analysis) as well as nonlinear associations among a single latent variable and a set of observed variables (by allowing a latent variable interaction among a variable and itself). The history of such methods dates back more than 50 years (e.g., McDonald 1962), but good estimation methods have not existed for high-dimensional applications until very recently (see Asparouhov & Muthén 2021).

A long-standing problem in structural equation models has been how researchers can estimate nonlinear associations in both structural and measurement models. Because latent variables are not directly observed, they must be estimated as a set of (co)variances and regression coefficients that constitute the key measurement and structural components of a structural equation model. Nowhere in this process are the underlying scores on latent variables themselves estimated. Instead, only their model-specified moment properties are estimated (e.g., variances, covariances, and structural regression coefficients). This makes it extremely difficult to estimate nonlinear models such as those involving statistical interactions, because such models are typically estimated by including product terms (e.g., $X*Z$) as predictor variables. This is not possible in SEM because underlying scores on latent variables do not exist in a dataset and are never actually estimated.

Various approaches have been taken to address this problem, with the most recent work in the organizational sciences focusing on a frequentist method referred to as latent moderated structural equations (LMS) (see Cheung et al. 2021, Su et al. 2019). For example, Toker & Biron (2012) used LMS to test the moderating effect of physical activity on the reciprocal relationship between burnout and depression. However effective such an approach may be, a much more tractable and efficient Bayesian alternative was recently proposed by Asparouhov & Muthén (2021). Although LMS is currently more common, it runs into computational difficulties in high dimensions that often lead to poor precision and convergence during estimation (for a discussion, see Preacher et al. 2016). The Bayesian alternative has already proven itself to be highly capable and extremely efficient, even with large samples and models with many latent variables. For example, Ozkok et al. (2021) recently showed how latent variable interactions could be formed and estimated with this Bayesian method in cross-lagged models to allow for interaction effects among within-person and/or between-person factors, including the cross-level interaction case. This would not have been possible using LMS.

Although both of these methods are implemented in the program Mplus, we recommend that researchers use the Bayesian approach in their research, in part because it can be easily applied even in very complex models, including multilevel SEM, but also because it provides more accurate results in almost all cases and it does not require that researchers make sacrifices related to precision and convergence (see Asparouhov & Muthén 2021). Conveniently, the method can also be applied to latent variable interactions involving a single latent variable, to allow for nonlinear relationships among a latent variable and its observed-variable indicators, as well as for nonlinear structural relationships with other latent variables. Such possibilities allow for numerous avenues for future research into how scales function and how latent variables may be related to each other—expanding the range of typical linear SEM to include nonlinear functional forms of measurement and structural relations. Indeed, it would be interesting to see how many past scale validation studies did not attempt to examine for nonlinear measurement and structural associations when conducting CFA (we suspect virtually all of them, given the historical difficulties with estimating such models). New nonlinear methods with latent interactions could then be used to evaluate any historically validated scales to test for nonlinear measurement and structural relationships that were previously missed.

Multilevel structural equation modeling. The next advanced modeling method we explore is MSEM (see Preacher et al. 2010, 2011, 2016). This recent approach vastly improves typical multilevel modeling methods that were popularized in the 1980s and 1990s, including hierarchical linear modeling. The problem with these historical methods is that although they enable a latent decomposition of an outcome variable into its level-specific components to account for sampling error (Raudenbush & Bryk 2002), they do not also do this for predictor variables and, even worse, they often lead to conflating level-specific effects when estimating the relationships between predictors and an outcome (see Zhang et al. 2009).

The solution to both problems is the automatic latent decomposition of observed variables into their level-specific components, which is the hallmark of MSEM—this method can also be understood as automatic latent group-mean centering of multilevel data (Hamaker & Muthén 2020). Conveniently, this decomposition is complemented by vast flexibility in the models that can be estimated, including multilevel CFA that enables investigating the multilevel measurement model characteristics of scales and constructs (Tay et al. 2014) in order to, among other things, estimate the reliability of a scale at different levels of analysis (Geldhof et al. 2013). The net result is a general approach to handling nested data structures that allows for the estimation of level-specific measurement and structural models, along with random slopes and even random variances at higher levels of analysis in order to estimate differences across groups, not only in terms of their averages but also the degree of within-group variation (see Lester et al. 2021, McNeish 2021).

In the organizational sciences, MSEM is becoming widely adopted due to its ability to address research questions that involve modeling teams and organizations. While preparing this review, we excluded 47 MSEM papers that were returned by our search (not including multilevel path analysis models). For example, D'Innocenzo et al. (2016) applied MSEM to test the moderating effect of unit-level empowerment on the relationship between individual psychological empowerment and individual performance. MSEM methods have also been applied to experience sampling method (ESM) data to model within- and between-person processes, as we further describe below. For example, Hülsheger et al. (2013) applied MSEM to ESM data in order to test the effect of mindfulness on surface acting and emotional exhaustion at the within- and between-person levels. Such applications are unique because they offer insights into phenomena and processes across multiple levels simultaneously. We anticipate substantial further updates of these methods and advocate for research designs that will allow estimating these more complex models.

Cross-lagged panel model and dynamic structural equation model. The next set of advancements for SEM applications involve models of longitudinal data, including single-level SEM specifications such as CLPMs that allow modeling panel data ranging from three to ten occasions of measurement ($T = 3$ to $T = 10$), to dynamic structural equation models that allow modeling intensive panel datasets ranging from roughly 10 to hundreds or even thousands of occasions of measurement. As an example of the former case, Schaubroeck et al. (2013) applied cross-lagged panel modeling to multiwave data and found that after accounting for the stability over time in cognition-based trust and affect-based trust, cognition-based trust influenced the development of affect-based trust for newcomers.

Notably, an important feature of hypothesis tests such as this has been the recognition in the organizational sciences and beyond that typical latent variable specifications allow accounting for stable factors in longitudinal data [e.g., individual differences in ESM data (see Allison 2009, Bollen & Brand 2010, Hamaker et al. 2015)]. This has led to a revolution in the ways that longitudinal panel data can be modeled in the organizational sciences. In the single-level case, recent work by numerous researchers has shown that classic applications of cross-lagged models—particularly those with only $T = 2$ occasions of measurement—confound stable traits with the within-person fluctuations that are often of interest for researchers (Zyphur et al. 2020a,b, 2021). In turn, this allows the construction of CLPMs with various types of fixed effects that account for these stable factors while estimating a variety of within-person lagged effects, including latent interactions and long-run effects of various types (see Ozkok et al. 2021, Shamsollahi et al. 2021; our use of the term fixed effects here also includes the random intercepts in the random intercept CLPM by Hamaker et al. 2015). By controlling for stable confounders, only the within-person fluctuations that reflect changes over time in a phenomenon can be isolated and used to infer predictive effects that are often referred to as Granger causality. Nevertheless, there are concerns and caveats that readers should be aware of regarding the use of both CLPMs and fixed effects variants that account for stable confounders (see Andersen 2021; Lucas 2022; Zyphur et al. 2020a,b).

On the other hand, with intensive longitudinal data, typical single-level structural equation models are not tractable. Therefore, MSEM approaches are needed that treat observations over time as nested within people to allow classic within-person and between-person variance decompositions. The new dynamic structural equation modeling (DSEM) framework allows for this, with at least $T = 10$ occasions of measurement being recommended (Schultzberg & Muthén 2018). Although historically such research designs have been rare in the organizational sciences, intensive longitudinal data are more often being collected from both archival sources as well as experience sampling/ecological momentary assessment studies (Gabriel et al. 2018). The logic of DSEM is relatively straightforward: Like the CLPM, stable factors are controlled when estimating within-person lagged associations (McNeish & Hamaker 2020). However, because the dynamic structural equation model is a multilevel model, it allows estimating random slopes and random variances for within-person model components, which can then be predicted at the between-person level (see Asparouhov et al. 2018, Hamaker et al. 2018, Zhou et al. 2019). By using these methods, researchers will be able to open up fundamentally new types of inquiry in the organizational sciences, which we wholeheartedly endorse.

Meta-analytic structural equation modeling. The final area of innovation we discuss is the advent and adoption of MA-SEM. This approach has seen substantial innovations in the past 20 years (Cheung & Chan 2005, Jak & Cheung 2020, Landis 2013), and there have been numerous recent applications in the organizational sciences (e.g., Bergh et al. 2016). The unique benefit of MA-SEM is that it combines typical meta-analytic procedures with the capabilities of SEM, allowing for structural model specification with meta-analytically derived covariance information as data inputs for the estimation of parameters and effect sizes. The result is a new way of making inferences about structural relationships that are derived from many primary studies (rather than only one). This helps address potential concerns of sampling error and generalizability of findings.

For example, Butts et al. (2013) applied MA-SEM to test a structural model of the effect of work-family support policies on job attitudes, which enabled them to go beyond bivariate relationships and single-study structural equation models to investigate the mediating role of variables such as work-to-family conflict and family-supportive organization perceptions in a collection of published studies. We anticipate many further developments in the MA-SEM space, and we believe, as other organizational researchers do (e.g., Landis 2013), that further developments such as methods for evaluating moderation in MA-SEM will offer new and exciting insights into organizational science.

CONCLUSION

In this review, we have provided an overview of a large number of published studies in organizational research that use SEM, describing their reporting practices and other attributes while making prescriptive judgments about how researchers should conduct and report structural equation models. Although SEM has been in use for decades within and outside of the organizational sciences, we find it troubling that there is still such a marked degree of underreporting of estimators, fit statistics, and other model attributes that are fundamental for evaluating SEM specifications and the results of estimation. In the future, we call on authors, reviewers, and editors to consider issues of replicability and the ability to critically evaluate a given SEM application, which is hindered without adequate information about the SEM specification, estimator, and resulting levels of model fit. In what follows, we make numerous observations about SEM and how we believe it can be put to good use to advance organizational science, discussing cross-validation, regularization, and problems with machine learning, as well as concerns about causality.

Cross-Validation, Regularization, and Problems with Machine Learning

One major flaw of almost all SEM applications is that model parameters are estimated using the same data that are used for evaluating model fit. This leads to well-known problems of overfitting, which reduces bias but increases variance in generalizations and therefore does not optimize the bias-variance tradeoffs that are present in our inferential practices (Schuler & Rose 2017). This problem is well known and has been the subject of substantial discussion in the machine learning and artificial intelligence literature (Belkin et al. 2019). The issue in the SEM literature is that the problem does not appear to be well understood and is virtually never discussed as a threat to inferences among the typical users of SEM. We see two paths forward for SEM to address this problem, which would need to be implemented in SEM software by its authors or by researchers using bespoke methods for their specific applications.

First, cross-validation methods can be automated so that one part of a sample (the training set) is used to estimate model parameters, and then the fit of the model is evaluated by using the estimate to make predictions about the remaining part of the sample (the test set). Advances of this classic holdout approach include *k*-fold methods and repeated holdouts that make better use of the information in a sample to evaluate the model fit (see Kim 2009). The underlying logic of such approaches in the SEM space would be based on a deterministic theory of relationships vis-à-vis the structural model specified. The hallmark of determinism is predictability, and therefore if the structural model specified is true to the world it is meant to represent, then the estimated parameters should allow predicting data that were not used for estimation. This is a reasonable expectation of SEM research. We believe that the time has come for SEM software to automate such methods so that researchers can better understand whether their model fits in data that remain unobserved when estimating model parameters.

The second approach to addressing the bias-variance tradeoff is regularization methods that automatically adjust parameter estimates so that bias (as misfit to the observed data) is increased, but variance (as nongeneralizability to new data) is reduced. Readers may be familiar with techniques, such as lasso regression, which make use of such estimators. Similar approaches are already used in multilevel modeling, including MSEM wherein empirical Bayes or shrinkage estimators are common and used by default to account for uncertainty in higher-level random effects (see Raudenbush & Bryk 2002). The basic idea with such approaches is that a penalization should be enacted to account for the tendency for maximum likelihood and similar estimators to overfit to an observed dataset, and we believe that this should be considered an appropriate way to estimate structural equation models in most situations (see Jacobucci et al. 2016, Li & Jacobucci 2021). The realization that researchers must make is that correcting for model complexity and a tendency to overfit a sample must be recognized outside of mere adjustments to fit indices (e.g., as the CFI, TLI, and other fit indices do). Instead, this adjustment must also be made to parameter estimates themselves. In our view, it is now time for SEM software to include (by default) estimators that take into account the tendency for most methods to overfit parameters to an observed dataset.

By taking advantage of such approaches to model estimation and inference, the SEM community will go a long way toward addressing some of SEM's disadvantages compared to machine learning and artificial intelligence methods. These methods are incredibly effective at prediction problems, including classification. However, they typically do not offer scientists what they want most: parsimonious causal explanations of the world, accompanied by estimates of uncertainty such as confidence intervals. Structural equation models have these qualities, which give them an advantage over the atheoretical alternatives that dominate other areas of quantitative research. Science is about more than prediction. Science is about explanation, but this does not mean that structural equation models and their associated estimators and inferential methods should not take advantage of developments from prediction-focused fields such as computer science.

Concerns About Causality

The ever-present concern when applying methods like SEM is that the estimated parameters do not offer a straightforward path to causal inference. This problem in the SEM space is typically conceived along classic lines as follows: Causality is a natural feature of the world, and research can uncover causal effects through controlled experiments, but such research designs make it hard to generalize to real-world situations; conversely, realistic observational designs allow for generalizable results, but this comes at the expense of strong causal inferences (for examples, see Scandura & Williams 2000). Resolving the inherent problem of more control versus more realism in research design (among other purportedly mutually exclusive study properties) has been jokingly called "dillemagic" by McGrath (1981, p. 181). However, in our view, this general approach to understanding causal inference vis-à-vis dilemmatic tradeoffs is outdated and needs to be revised when thinking about causality and SEM for multiple reasons.

The first is the possibility of enhancing causal inferences through methods such as instrumental variable analysis, wherein observational designs with desirable properties (e.g., unobtrusiveness, large samples, realistic settings) are combined with a randomization device to enhance causal inference (Antonakis et al. 2010). Related methods, including pre- and postnatural experiments (which can be seen as types of instrumental variable analysis) and propensity score methods can at least partially address some of the historical concerns about the observational research designs that are typically used to collect data in much SEM research.

However, second and more importantly, recent rigorous investigations of what causality is, and how it should be understood from both scientific and policy perspectives, call into question the simplistic notions of causal inference on which concerns about research design have historically been based. Specifically, exemplary work by Cartwright and colleagues (e.g., Cartwright & Hardie 2012, Deaton & Cartwright 2018) has critically interrogated the very notion of evidence (e.g., Cartwright 2017, Cartwright et al. 2009), and no strong theory of evidence appears to be in use in many social scientific and health sciences fields. For example, evidence hierarchies are very common, but not only are these not a theory of evidence, they often serve to distract us from the hard work of developing a deep understanding of what should count as evidence and by how much in a given context for a given purpose. It is this type of consideration—what should count as evidence here in this context for this specific purpose—that should ground conceptualizations of evidence and the validity of causal inferences, rather than the general and formulaic approaches that have historically been used to understand SEM research and its associated conclusions (e.g., via conceptions of internal or external validity).

The truth about causality is that although it may be a natural feature of the world, this does not mean that causal inferences can or should be decontextualized from their domains of application. For example, an RCT may be able to establish that a causal effect occurred in a given study, but that does not mean the scientific finding is relevant to a real-world situation in which the study result will be applied. As Cartwright & Hardie (2012) describe:

> You want evidence that a policy will work here, where you are. Randomized controlled trials do not tell you that. They do not even tell you that a policy works. What they tell you is that a policy worked there, where the trial was carried out... Our argument is that the changes in tense—from "worked" to "work" to "will work"—are not just a matter of grammatical detail. To move from one to the other requires hard intellectual and practical effort. The fact that it worked there is indeed fact. But for that fact to be evidence that it will work here, it needs to be relevant to that conclusion. To make RCTs relevant you need a lot more information. (p. ix)

Conversely, we can make the reverse inference about observational data, which are often used to estimate structural equation models: Although these may not involve forms of controlled experimentation, this does not mean that they are irrelevant for understanding a given real-world situation and how to intervene in it. From this perspective, an RCT and an observational SEM study may be equally relevant or irrelevant for practical action in the world, because practical action is always carried out in a given context for a given purpose. Therefore, we recommend that future work on the issue of causality and inference reconsiders both the critique of observational structural equation models and the lauding of RCTs without deeper considerations of exactly what the concept of evidence is meant to secure from a practical perspective (for a discussion, see Diener et al. 2022).

Organizational researchers should not get stuck in the trap of assuming that their goal is to produce causal inferences that are "true" in an abstract sense. The generalizability of a causal inference is found when it is enacted in a real-world situation and generates one or desired outcomes after a study has been published and its results are used to guide action (Zyphur & Pierides 2017, 2020a). To this end, future work on SEM and causality should consider how structural equation models can be applied so that their results can be more easily put into practice to generate beneficial real-world outcomes for individuals and society. This starts with good reporting practices as we have described, but it also means considering what kinds of problems and contexts a given SEM study's results could be profitably applied to, and then describing why and how, rather than using somewhat simplistic and antiquated notions of validity or evidence that may distract researchers and practitioners from what makes the research useful. SEM is useful because it is a very generalizable method, offering many ways to represent and study real-world phenomena. By better linking SEM data and models to specific potential domains of real-world application, organizational researchers can maximize the potential benefits and usefulness of their research using the diversity of modeling and estimation approaches we have described here, all of which fall within the incredibly powerful modeling framework that is SEM.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

We would like to thank the Editorial Committee and Production staff for their very helpful comments and assistance revising the manuscript.

LITERATURE CITED

Allison PD. 2009. *Fixed Effects Regression Models*. Thousand Oaks, CA: SAGE

Ancona DG, Goodman PS, Lawrence BS, Tushman ML. 2001. Time: a new research lens. *Acad. Manag. Rev.* 26:645–83

Andersen HK. 2021. Equivalent approaches to dealing with unobserved heterogeneity in cross-lagged panel models? Investigating the benefits and drawbacks of the latent curve model with structured residuals and the random intercept cross-lagged panel model. *Psychol. Methods.* In press. **https://doi.org/10.1037/met0000285**

Anderson JC, Gerbing DW. 1988. Structural equation modeling in practice: a review and recommended two-step approach. *Psychol. Bull.* 103(3):411–23

Antonakis J, Bendahan S, Jacquart P, Lalive R. 2010. On making causal claims: a review and recommendations. *Leadersh. Q.* 21:1086–1120

Asparouhov T, Hamaker EL, Muthén B. 2018. Dynamic structural equation models. *Struct. Equ. Model.* 25(3):359–88

Asparouhov T, Muthén B. 2010. *Plausible values for latent variables using Mplus*. Work. Pap., Mplus, Los Angel. **https://www.statmodel.com/download/Plausible.pdf**

Asparouhov T, Muthén B. 2021. Bayesian estimation of single and multilevel models with latent variable interactions. *Struct. Equ. Model.* 28(2):314–28

Baum JAC, Haveman HA. 2020. Editors' comments: the future of organizational theory. *Acad. Manag. Rev.* 45:268–72

Belkin M, Hsu D, Ma S, Mandal S. 2019. Reconciling modern machine-learning practice and the classical bias-variance trade-off. *PNAS* 116(32):15849–54

Bentler PM, Bonett DG. 1980. Significance tests and goodness of fit in the analysis of covariance structures. *Psychol. Bull.* 88:588–606

Bentler PM, Satorra A. 2010. Testing model nesting and equivalence. *Psychol. Methods* 15:111–23
Bergh DD, Aguinis H, Heavey C, Ketchen DJ, Boyd BK, et al. 2016. Using meta-analytic structural equation modeling to advance strategic management research: guidelines and an empirical illustration via the strategic leadership-performance relationship. *Strateg. Manag. J.* 37(3):477–97
Binning JF, Barrett GV. 1989. Validity of personnel decisions: a conceptual analysis of the inferential and evidential bases. *J. Appl. Psychol.* 74:478–94
Bollen KA. 2002. Latent variables in psychology and the social sciences. *Annu. Rev. Psychol.* 53:605–34
Bollen KA, Brand JE. 2010. A general panel model with random and fixed effects: a structural equations approach. *Soc. Forces* 89(1):1–34
Bollen KA, Long JS, eds. 1993. *Testing Structural Equation Models*. Thousand Oaks, CA: SAGE
Borsboom D, Mellenbergh GJ, van Heerden J. 2003. The theoretical status of latent variables. *Psychol. Rev.* 110:203–19
Browne MW, Cudeck R. 1993. Alternative ways of assessing model fit. *Sociol. Methods Res.* 21(2):230–58
Bullock HE, Harlow LL, Mulaik SA. 1994. Causation issues in structural equation modeling research. *Struct. Equ. Model.* 1(3):253–67
Butts MM, Casper WJ, Yang TS. 2013. How important are work–family support policies? A meta-analytic investigation of their effects on employee outcomes. *J. Appl. Psychol.* 98(1):1–25
Byrne BM, Shavelson RJ, Muthén B. 1989. Testing for the equivalence of factor covariances and mean structures: the issue of partial measurement invariance. *Psychol. Bull.* 105:456–66
Cartwright N. 2017. Single case causes: what is evidence and why. In *Philosophy of Science in Practice*, ed. H-C Chao, J Reiss, pp. 11–24. Cham, Switz.: Springer
Cartwright N, Goldfinch A, Howick J. 2009. Evidence-based policy: Where is our theory of evidence? *J. Child. Serv.* 4(4):6–14
Cartwright N, Hardie J. 2012. *Evidence-Based Policy: A Practical Guide to Doing It Better*. Oxford, UK: Oxford Univ. Press
Cheung GW, Cooper-Thomas HD, Lau RS, Wang LC. 2021. Testing moderation in business and psychological studies with latent moderated structural equations. *J. Bus. Psychol.* 36:1009–33
Cheung MWL, Chan W. 2005. Meta-analytic structural equation modeling: a two-stage approach. *Psychol. Methods* 10(1):40–64
Connelly BS, McAbee ST, Oh I-S, Jung Y, Jung C-W. 2022. A multirater perspective on personality and performance: an empirical examination of the trait–reputation–identity model. *J. Appl. Psychol.* 107(8):1352–68
Cortina JM, Green JP, Keeler KR, Vandenberg RJ. 2017. Degrees of freedom in SEM: Are we testing the models we claim to test? *Organ. Res. Methods* 20(3):350–78
D'Innocenzo L, Luciano MM, Mathieu JE, Maynard MT, Chen G. 2016. Empowered to perform: a multilevel investigation of the influence of empowerment on performance in hospital units. *Acad. Manag. J.* 59(4):1290–1307
Davidov E, Muthen B, Schmidt P. 2018. Measurement invariance in cross-national studies: challenging traditional approaches and evaluating new ones. *Sociol. Methods Res.* 47:631–36
Deaton A, Cartwright N. 2018. Understanding and misunderstanding randomized controlled trials. *Soc. Sci. Med.* 210:2–21
Denis D, Legerski J. 2006. Causal modeling and the origins of path analysis. *Theory Sci.* 7(1):2–10
Diener E, Northcott R, Zyphur MJ, West SG. 2022. Beyond experiments. *Perspect. Psychol. Sci.* 17(4):1101–19
Edwards JR. 2010. The fallacy of formative measurement. *Organ. Res. Methods* 14(2):370–388
Edwards JR, Bagozzi RP. 2000. On the nature and direction of relationships between constructs and measures. *Psychol. Methods* 5:155–74
Ferris DL, Yan M, Lim VK, Chen Y, Fatimah S. 2016. An approach–avoidance framework of workplace aggression. *Acad. Manag. J.* 59:1777–1800
Foulk TA, Woolum A, Erez A. 2016. Catching rudeness is like catching a cold: the contagion effects of low-intensity negative behaviors. *J. Appl. Psychol.* 101(1):50–67
Frone MR, Russell M, Cooper ML. 1992. Antecedents and outcomes of work-family conflict: testing a model of the work-family interface. *J. Appl. Psychol.* 77(1):65–78

Frone MR, Russell M, Cooper ML. 1994. Relationship between job and family satisfaction: Causal or noncausal covariation? *J. Manag.* 20(3):565–79

Fulmer CA, Ostroff C. 2017. Trust in direct leaders and top leaders: a trickle-up model. *J. Appl. Psychol.* 102(4):648–57

Gabriel AS, Podsakoff NP, Beal DJ, Scott BA, Sonnentag S, et al. 2018. Experience sampling methods. *Organ. Res. Methods.* 22(4):969–1006

Geldhof GJ, Preacher KJ, Zyphur MJ. 2013. Reliability estimation in a multilevel confirmatory factor analysis framework. *Psychol. Methods.* 19(1):72–91

Grace JB. 2021. Instrumental variable methods in structural equation models. *Methods Ecol. Evol.* 12(7):1148–57

Hamaker EL, Asparouhov T, Brose A, Schmiedek F, Muthén B. 2018. At the frontiers of modeling intensive longitudinal data: dynamic structural equation models for the affective measurements from the COGITO study. *Multivar. Behav. Res.* 53(6):820–41

Hamaker EL, Kuiper RM, Grasman RP. 2015. A critique of the cross-lagged panel model. *Psychol. Methods* 20(1):102–16

Hamaker EL, Muthén B. 2020. The fixed versus random effects debate and how it relates to centering in multilevel modeling. *Psychol. Methods* 25(3):365–79

Hayes T, Usami S. 2020. Factor score regression in the presence of correlated unique factors. *Educ. Psychol. Meas.* 80(1):5–40

Hershcovis MS, Bhatnagar N. 2017. When fellow customers behave badly: witness reactions to employee mistreatment by customers. *J. Appl. Psychol.* 102(11):1528–44

Hoyle RH, Smith GT. 1994. Formulating clinical research hypotheses as structural equation models: a conceptual overview. *J. Consult. Clin. Psychol.* 62:429–40

Hu LT, Bentler PM. 1999. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct. Equ. Model.* 6:1–55

Hülsheger UR, Alberts HJEM, Feinholdt A, Lang JWB. 2013. Benefits of mindfulness at work: the role of mindfulness in emotion regulation, emotional exhaustion, and job satisfaction. *J. Appl. Psychol.* 98:310–25

Jacobucci R, Grimm KJ, McArdle JJ. 2016. Regularized structural equation modeling. *Struct. Equ. Model.* 23(4):555–66

Jak S, Cheung MWL. 2020. Meta-analytic structural equation modeling with moderating effects on SEM parameters. *Psychol. Methods* 25(4):430–55

Jorgensen TD, Pornprasertmanit S, Schoemann AM, Rosseel Y. 2022. semTools: useful tools for structural equation modeling. R package version 0.5–6. *semTools.* **https://CRAN.R-project.org/package=semTools**

Kim J-H. 2009. Estimating classification error rate: repeated cross-validation, repeated hold-out and bootstrap. *Comput. Stat. Data Anal.* 53(11):3735–45

Kline R. 2004. *Principles and Practice of Structural Equation Modeling*. New York: Guilford Press. 2nd ed.

Kunze F, Raes AML, Bruch H. 2015. It matters how old you feel: antecedents and performance consequences of average relative subjective age in organizations. *J. Appl. Psychol.* 100(5):1511–26

Kuykendall L, Craig L, Tay L. 2019. Work-contingent self-esteem: a boon or bane for worker well-being? *J. Organ. Behav.* 41(1):1–16

Landis RS. 2013. Successfully combining meta-analysis and structural equation modeling: recommendations and strategies. *J. Bus. Psychol.* 28(3):251–61

Lester HF, Cullen-Lester KL, Walters RW. 2021. From nuisance to novel research questions: using multilevel models to predict heterogeneous variances. *Organ. Res. Methods* 24(2):342–88

Li X, Jacobucci R. 2021. Regularized structural equation modeling with stability selection. *Psychol. Methods* 27(4):497–518

Livne-Ofer E, Coyle-Shapiro JA, Pearce JL. 2019. Eyes wide open: perceived exploitation and its consequences. *Acad. Manag. J.* 62(6):1989–2018

Loi TI, Kuhn KM, Sahaym A, Betterfield KD, Tripp TM. 2020. From helping hands to harmful acts: when and how employee volunteering promotes workplace deviance. *J. Appl. Psychol.* 105(9):944–58

Lucas R. 2022. It's time to abandon the cross-lagged panel model. PsyArXiv pkec7. **https://psyarxiv.com/pkec7/**

MacCallum RC, Austin JT. 2000. Applications of structural equation modeling in psychological research. *Annu. Rev. Psychol.* 51:201–26

MacCallum RC, Wegener DT, Uchino BN, Fabrigar LR. 1993. The problem of equivalent models in applications of covariance structure analysis. *Psychol. Bull.* 114:185–99

MacKenzie SB, Podsakoff PM, Jarvis CB. 2005. The problem of measurement model misspecification in behavioral and organizational research and some recommended solutions. *J. Appl. Psychol.* 90(4):710–30

Maydeu-Olivares A, Shi D, Fairchild AJ. 2020. Estimating causal effects in linear regression models with observational data: the instrumental variables regression model. *Psychol. Methods* 25(2):243–58

McCarthy JM, Trougakos JP, Cheng BH. 2016. Are anxious workers less productive workers? It depends on the quality of social exchange. *J. Appl. Psychol.* 101(2):279–91

McDonald RP. 1962. A general approach to nonlinear factor analysis. *Psychometrika* 27:397–415

McDonald RP. 2010. Structural models and the art of approximation. *Perspect. Psychol. Sci.* 5(6):675–86

McDonald RP, Ho M-HR. 2002. Principles and practice in reporting structural equation analyses. *Psychol. Methods* 7(1):64–82

McGrath JE. 1981. The study of research choices and dilemmas. *Am. Behav. Sci.* 25:179–210

McNeish D. 2021. Specifying location-scale models for heterogeneous variances as multilevel SEMs. *Organ. Res. Methods* 24:630–53

McNeish D, Hamaker EL. 2020. A primer on two-level dynamic structural equation models for intensive longitudinal data in Mplus. *Psychol. Methods* 25(5):610–35

McNeish D, Wolf MG. 2021. Dynamic fit index cutoffs for confirmatory factor analysis models. *Psychol. Methods*. In press. **https://doi.org/10.1037/met0000425**

Muthén B, Asparouhov T. 2015. Causal effects in mediation modeling: an introduction with applications to latent variables. *Struct. Equ. Model.* 22:12–23

Muthén B, Asparouhov T. 2017. Recent methods for the study of measurement invariance with many groups. *Sociol. Methods Res.* 47(4):637–64

Muthén BO, Muthén LK, Asparouhov T. 2017. *Regression and Mediation Analysis Using Mplus*. Los Angeles: Mplus

Nye CD, Brummel BJ, Drasgow F. 2014. Understanding sexual harassment using aggregate construct models. *J. Appl. Psychol.* 99(6):1204–21

O'Boyle EH Jr., Williams LJ. 2011. Decomposing model fit: measurement versus theory in organizational research using latent variables. *J. Appl. Psychol.* 96(1):1–12

Ozkok O, Vaulont MJ, Zyphur MJ, Zhang Z, Preacher KJ, et al. 2021. Interaction effects in cross-lagged panel models: SEM with latent interactions applied to work-family conflict, job satisfaction, and gender. *Organ. Res. Methods.* 25(4):673–715

Pearl J. 2010. The foundations of causal inference. *Sociol. Methodol.* 40:75–149

Petter S, Straub D, Rai A. 2007. Specifying formative constructs in information systems research. *MIS Q.* 31(4):623–56

Ployhart RE, Oswald FL. 2004. Application of mean and covariance structures analysis: integrating correlational and experimental approaches. *Organ. Res. Methods* 7:27–65

Preacher KJ, Zhang Z, Zyphur MJ. 2010. A general multilevel SEM framework for assessing multilevel mediation. *Psychol. Methods* 15:209–33

Preacher KJ, Zhang Z, Zyphur MJ. 2011. Alternative methods for assessing mediation in multilevel data: the advantages of multilevel SEM. *Struct. Equ. Model.* 18(2):161–82

Preacher KJ, Zhang Z, Zyphur MJ. 2016. Multilevel structural equation models for assessing moderation within and across levels of analysis. *Psychol. Methods* 21(2):189–205

Priesemuth M, Schminke M, Ambrose ML, Folger R. 2014. Abusive supervision climate: a multiple-mediation model of its impact on group outcomes. *Acad. Manag. J.* 57(5):1513–34

Quigley TJ, Hubbard TD, Ward A, Graffin SD. 2020. Unintended consequences: information releases and CEO stock option grants. *Acad. Manag. J.* 63:155–80

Raudenbush SW, Bryk AS. 2002. *Hierarchical Linear Models: Applications and Data Analysis Methods.* Thousand Oaks, CA: SAGE. 2nd ed.

Raykov T, Tomer A, Nesselroade JR. 1991. Reporting structural equation modeling results in *Psychology and Aging*: some proposed guidelines. *Psychol. Aging* 6(4):499–503

Rosseel Y, Loh WW. 2021. *A "Structural-After-Measurement" (SAM) approach to SEM*. OSF: Center for Open Science's Open Source Web Application, Charlottesville, VA, retrieved August 15, 2022. **https://osf.io/pekbm/**

Russell DW, Kahn JH, Spoth R, Altmaier EM. 1998. Analyzing data for experimental studies: a latent variable structural equation modeling approach. *J. Couns. Psychol.* 45:18–29

Saylors R, Trafimow D. 2020. Why the increasing use of complex causal models is a problem: on the danger sophisticated theoretical narratives pose to truth. *Organ. Res. Methods* 24(3):616–29

Scandura TA, Williams EA. 2000. Research methodology in management: current practices, trends, and implications for future research. *Acad. Manag. J.* 43(6):1248–64

Schaubroeck JM, Peng AC, Hannah ST. 2013. Developing trust with peers and leaders: impacts on organizational identification and performance during entry. *Acad. Manag. J.* 56(4):1148–68

Schreiber JB, Nora A, Stage FK, Barlow EA, King J. 2010. Reporting structural equation modeling and confirmatory factor analysis results: a review. *J. Educ. Res.* 99(6):323–38

Schuler MS, Rose S. 2017. Targeted maximum likelihood estimation for causal inference in observational studies. *Am. J. Epidemiol.* 185(1):65–73

Schultzberg M, Muthén B. 2018. Number of subjects and time points needed for multilevel time-series analysis: a simulation study of dynamic structural equation modeling. *Struct. Equ. Model.* 25(4):495–515

Shamsollahi A, Zyphur MJ, Ozkok O. 2021. Long-run effects in dynamic systems: new tools for cross-lagged panel models. *Organ. Res. Methods.* 25(3):435–58

Skrondal A, Laake P. 2001. Regression among factor scores. *Psychometrika* 66:563–75

Solomon BC, Nikolaev BN, Shepherd DA. 2021. Does educational attainment promote job satisfaction? The bittersweet trade-offs between job resources, demands, and stress. *J. Appl. Psychol.* 10(7):1227–41

Su R, Zhang Q, Liu Y, Tay L. 2019. Modeling congruence in organizational research with latent moderated structural equations. *J. Appl. Psychol.* 104(11):1404–33

Tarka P. 2018. An overview of structural equation modeling: its beginnings, historical development, usefulness and controversies in the social sciences. *Qual. Quant.* 52:313–54

Tay L, Meade AW, Cao M. 2015. An overview and practical guide to IRT measurement equivalence analysis. *Organ. Res. Methods* 18:3–46

Tay L, Woo SE, Vermunt JK. 2014. A conceptual and methodological framework for psychometric isomorphism: validation of multilevel construct measures. *Organ. Res. Methods* 17:77–106

Toker S, Biron M. 2012. Job burnout and depression: unraveling their temporal relationship and considering the role of physical activity. *J. Appl. Psychol.* 97(3):699–710

Vandenberg RJ, Grelle DM. 2009. Alternative model specifications in structural equation modeling. In *Statistical and Methodological Myths and Urban Legends*, ed. CE Lance, RJ Vandenberg, pp. 165–91. London: Routledge

Vandenberg RJ, Lance CE. 2000. A review and synthesis of the measurement invariance literature: suggestions, practices, and recommendations for organizational research. *Organ. Res. Methods* 3:4–70

Viswesvaran C, Ones DS. 1995. Theory testing: combining psychometric meta-analysis and structural equations modeling. *Pers. Psychol.* 48:865–85

Vrieze SI. 2012. Model selection and psychological theory: a discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). *Psychol. Methods* 17(2):228–43

Wanberg CR, Csillag B, Douglass RP, Zhou L, Pollard MS. 2020. Socioeconomic status and well-being during COVID-19: a resource-based examination. *J. Appl. Psychol.* 105(12):1382–96

Weston R, Gore PA. 2016. A brief guide to structural equation modeling. *Couns. Psychol.* 34(5):719–51

Williams LJ, Vandenberg RJ, Edwards JR. 2009. Structural equation modeling in management research: a guide for improved analysis. *Acad. Manag. Ann.* 3:543–604

Wo DXH, Ambrose ML, Schminke M. 2015. What drives trickle-down effects? A test of multiple mediation processes. *Acad. Manag. J.* 58(6):1848–68

Zhang XA, Liao H, Li N, Colbert AE. 2020. Playing it safe for my family: exploring the dual effects of family motivation on employee productivity and creativity. *Acad. Manag. J.* 63:1923–50

Zhang Z, Zyphur MJ, Preacher KJ. 2009. Testing multilevel mediation using hierarchical linear models: problems and solutions. *Organ. Res. Methods* 12:695–719

Zhou L, Wang M, Zhang Z. 2019. Intensive longitudinal data analyses with dynamic structural equation modeling. *Organ. Res. Methods* 24(2):219–50

Zyphur MJ. 2009. When mindsets collide: switching analytical mindsets to advance organization science. *Acad. Manag. Rev.* 34(4):677–88

Zyphur MJ, Allison PD, Tay L, Voelkle MC, Preacher KJ, et al. 2020a. From data to causes I: building a general cross-lagged panel model (GCLM). *Organ. Res. Methods* 23:651–87

Zyphur MJ, Hamaker EL, Tay L, Voelkle M, Preacher KJ, et al. 2021. From data to causes III: Bayesian priors for general cross-lagged panel models (GCLM). *Front. Psychol.* 12:612251

Zyphur MJ, Pierides DC. 2017. Is quantitative research ethical? Tools for ethically practicing, evaluating, and using quantitative research. *J. Bus. Ethics* 143:1–16

Zyphur MJ, Pierides DC. 2020a. Making quantitative research work: from positivist dogma to actual social scientific inquiry. *J. Bus. Ethics* 167:49–62

Zyphur MJ, Pierides DC. 2020b. Statistics and probability have always been value-laden: an historical ontology of quantitative research methods. *J. Bus. Ethics* 167:1–18

Zyphur MJ, Pierides DC, Roffe J. 2015. Measurement and statistics in ‘organization science’: philosophical, sociological, and historical perspectives. In *The Routledge Companion to Philosophy in Organization Studies*, ed. R Mir, H Willmott, M Greenwood, pp. 474–82. London: Routledge

Zyphur MJ, Voelkle MC, Tay L, Allison PD, Preacher KJ, et al. 2020b. From data to causes II: comparing approaches to panel data analysis. *Organ. Res. Methods* 23:688–716

Zyphur MJ, Zhang Z, Preacher KJ, Bird L. 2019. Moderated mediation in multilevel structural equation models: decomposing effects of race on math achievement within versus between high schools in the United States. In *The Handbook of Multilevel Theory*, ed. SE Humphrey, JM LeBreton, pp. 473–94. Am. Psychol. Assoc.