Give Me Attitude: Making Smart Use of Structural Equation Modeling and Other Tools When Studying Survey Data

Nathan Favero
American University
Aarhus University

School of Public Affairs
4400 Massachusetts Ave. NW
Washington, DC 20016
favero@american.edu

Prepared for presentation at the annual meeting of the Association for Education Finance & Policy, Baltimore, MD, March 14-16, 2024. I would like to thank researchers at the Roskilde School of Governance for their comments on this project.

Give Me Attitude: Making Smart Use of Structural Equation Modeling and Other Tools When Studying Survey Data

Abstract: Many researchers were trained primarily in the tools of either econometric-style regression or psychology-inspired structural equation modeling (SEM). Given that quantitative research in public policy and administration regularly draws on both traditions of modeling, researchers frequently find themselves in the position of reading and evaluating studies that utilize modeling tools they don't understand well. This study aims to address this problem through a detailed comparison of each set of tools. I argue that in many cases, econometric and SEM tools yield similar substantive conclusions, though they also tend to be used to answer slightly different questions. Econometric tools fit more naturally with research designs intended to probe issues of causality, while SEM is better suited to generating detailed descriptions of patterns of associations among several variables. SEM also encourages researchers to think carefully about both mediation and measurement, and it offers greater flexibility in correcting for measurement error. At the same time, measurement error in the real world is typically even more complicated than what SEM can account for. Econometrics tools offer more options for dealing with measurement error than many researchers realize, and econometric models offer more flexibility in terms of the number and type of variables that can be realistically analyzed. The study concludes with a practical set of guidelines for analyzing hard-to-measure constructs—guidelines which can be used by researchers regardless of which disciplinary tradition of statistical modeling they draw on.

The study of attitudes (and other variables that cannot be measured in well-defined units) has long occupied an important space in quantitative social science research. Because such variables defy objective measurement, they are necessarily subject to notable measurement error. Within the public policy and administration literatures, there is a well-established interest in understanding a variety of hard-to-quantify variables, despite the difficulties posed by measurement issues. Examples include studies of employee motivations, job satisfaction, leadership behaviors, and citizen/client satisfaction. Among these hard-to-measure variables are several individual-level variables that are difficult to measure in any way except through self-reports. The lack of alternatives is often due to a limited visibility by others into one's own beliefs or psychological states. As such, we see that variables like attitudes, emotions, perceptions, intentions, and personal traits (such as personality) are measured by quantitative researchers almost exclusively through self-reports in survey questionnaires.

Over the past fifteen years, the field of public administration in particular has come to hold great skepticism toward studies employing common-source survey measures, due to concerns surrounding common source bias (Meier & O'Toole 2012). Yet several scholars have noted the lack of any practical quantitative alternative to relying on common-source survey measures when studying the link between two variables that both require measurement through self-reporting (at least in the context of observational studies) (George & Pandey 2017). For example, observational studies of the link between two attitudes can hardly avoid a common source measurement approach.

What are we to do then, as applied researchers, when facing a such research questions? So far, the public administration (PA) literature offers few answers. Surely, we should not just avoid quantitative study of such research questions. Some authors (including myself in the past) have occasionally held up structural equation models (SEM) as a potentially transformative method for analyzing data from a common survey source. But this approach has not thus far become predominant in PA, and many PA scholars lack much familiarity with SEM tools. Experimental (and quasi-experimental) research provides one promising path forward for studying linkages among hard-to-measure variables, since it is often possible to plausibly manipulate a hard-to-measure independent variable. At the same time, such designs often yield results with questionable external validity. Many experimental studies involve interventions that very temporarily alter the valence or salience of an attitude (Marvel 2016), or else they rely on fictitious description of a hypothetical world that the respondent must try to inhabit. Such designs can help to firmly establish the psychological underpinnings of important associations among variables. But a comprehensive understanding of social phenomena arguably requires that such experimental work can be paired with high-quality descriptive or observational studies that can enhance our confidence in external validity.

This article sets out to address two questions of relevance to many policy and administration researchers with interests in attitudes or other hard-to-measure variables: (1) How can one produce high-quality observational/descriptive work from a common survey source while accounting for the critiques leveled against reliance on common-source data? (2) What role should SEM fill in our analytical toolbox, and how can SEM analysis be critically assessed by policy and administration researchers? For the second question, I start from the assumption than while many researchers in our fields lack formal training in SEM, they sometimes find themselves reading or even reviewing studies that employ SEM. Perhaps, some readers also wonder whether they should start using SEM in their own studies. In this article, I will not attempt to equip readers to conduct rigorous SEM analyses themselves. Instead, I hope better enable readers to critically interpret and evaluate SEM analyses conducted by others. For readers who are interested in going deeper, there are abundant introductory treatments for how to conduct SEM available elsewhere; and maybe this article will serve to better inform readers about the potential utility of getting tooled up in this approach.

Three themes will underlie my arguments regarding the study of attitudes. First, SEM and traditional regression approaches are not so different. In my estimation, the differences between these approaches have been exaggerated in the literature. It is often the case that a similar set of questions can be plausibly answered with either set of tools, often with similar results. Second, simplest is often best. When multiple analytical approaches are possible, it is often best to go with the simpler approach because its interpretation (and its limitations) will be better understood by both the researcher and the researcher's audience. As the old adage says, all models are wrong, but some are useful. A useful model must be one that is understood. Third, measurement is key. Better measurement will produce better data, which will enable more reliable results. Often, the best measures are highly imperfect. Still, the weaknesses of measures should be carefully evaluated and borne in mind when interpreting results.

Challenges in Studying Attitudes

Researchers conducting quantitative studies of attitudes face two common challenges that will inform this discussion of methods.

First, datasets containing measures of attitudes often contain fairly high levels of multicollinearity. Even relatively modest multicollinearity among predictors can make regression results highly sensitive to small modeling changes. Decisions about what to include or omit from a regression can dramatically affect the results. Thus, it is very important to carefully consider model specification when working with attitudinal data. Particular attention should be paid to the possibility of mediation relationships, since controlling for mediating variables will generally bias one's results. Thus, running regression analysis with a set of attitudinal predictors generally requires that the researcher make fairly strict (perhaps implicit) assumptions about the sequence of causality—specifically, whether given variables are potential mediators (and thus should be omitted from the regression) or might be related to the key predictor and dependent variable through causal paths that require the inclusion of them as control variables.

In many cases, it may be difficult or impossible to precisely identify causal pathways among attitudinal variables. Many attitudinal variables likely exhibit reciprocal causality, making it difficult to delineate a single clear pathway of mediation. Attitudinal variables may also be conceptually overlapping to various degrees, making it impossible to conceptualize them as fully distinctive. Before conducting a complicated regression-based analysis with several attitudinal variables, one should perhaps first consider the question of whether their goal is to create a model of distinct variables linked by clear causal pathways. In some cases, it may be better to pursue the objective of mapping out the degrees of linkage among a set of intertwined attitudes, assessing the extent to which certain pairs of attitudes are more proximal or distal. If the research objectives are best realized through the latter approach, a simple set of bivariate correlations may be the best approach to empirically analyzing the data.

A second challenge when working with attitudinal data is the presence of substantial measurement error. Arguably the most defining feature of the psychometric tradition is its focus on measurement error, which is no surprise given the centrality of attitudes and other hard-to-quantify constructs to the field of psychology. Perhaps the dominant approach to quantitatively measuring such types of variables is to rely on surveys employing Likert scale survey response options, which generally require respondents to make highly subjective judgements when answering survey questions since the anchors associated with response options are typically poorly defined (e.g., "somewhat agree").

When considering the effects of measurement error, it is often useful to distinguish among some of the various sources of such error. For purposes of this study, I offer a simple set of equations to distinguish conceptually among three categories of measurement error. Suppose I measure teacher satisfaction with a three-item scale, consisting of the following three items:

$$satis1 = satis + e_{satis1} + e_{satis} + [wicked\_error]$$

$$satis2 = satis + e_{satis2} + e_{satis} + [wicked\_error]$$

$$satis3 = satis + e_{satis3} + e_{satis} + [wicked\_error]$$

Each observed variable $satis$j partially reflects the unobserved latent variable $satis$, representing the true value of teacher satisfaction. Additionally, each variable has a random error term ($e_{satis\text{j}}$) that is specific to that item j. Classical test theory generally considers only this type of error, which will usually be relatively straightforward to deal with so long as it is uncorrelated (or only weakly correlated) with the value of the latent variable and with other error terms. Given that such error is what a Cronbach's alpha is used assess, I refer to this type of error as Cronbach errors. In addition to the item-specific error, I consider the possibility that there scale-specific error (e_{satis}), which represents the case when there is measurement error which affects all items on the scale but is unrelated to the values or measurement errors of other constructs besides satisfaction. And finally, there may be what I call wicked errors of various sorts, such as common method variance that affects the measurement of multiple constructs in a study.

Given the prominence of Cronbach's alpha as the dominant tool for initial assessments of the reliability of attitudinal measures throughout the social sciences, it is perhaps unsurprising that researchers sometimes appear to emphasize minimizing Cronbach errors at the expensive of other types of measurement error. A prime example is engineering survey scales to be needlessly repetitive, yielding extremely similar responses to each item and thus minimizing the amount of item-specific variance. For example, Grant (2008) adopts the following measure:

> An introductory question asked, "Why are you motivated to do your work?"…. The four items for prosocial motivation were "Because I care about benefiting others through my work," "Because I want to help others through my work," "Because I want to have positive impact on others," and "Because it is important to me to do good for others through my work" (α = .90). (51)

The repetitive language makes the items extremely similar to one another, so it is no surprise that a high level of internal consistency is found, yielding a large Cronbach's alpha despite the small number of scale items. This reflects the small variance in item-specific measurement errors (Cronbach errors). Nonetheless, we would not expect that having extremely repetitive items leads to a more precise measure (less overall measurement error) of the actual construct (prosocial motivation) than would be obtained if similarly valid but more distinctly-worded items were used. Instead, the repetitive wording likely serves to minimize Cronbach errors at the expense of greater scale-specific error. Put differently, to the extent that these items will not perfectly capture the construct of prosocial motivation, the four items are likely erring in the same direction. With a more diversely-worded set of items, it would be more likely that the measurement errors of the items would be more independent (conceptualized here as having a larger item-level error and a smaller scale-level error). Despite the smaller Cronbach's alpha that we would expect from the more diversely-worded set of items, the index would likely be more useful since Cronbach errors are generally easier to correct for than scale-level errors.

Both Cronbach errors ($e_{satis\mathrm{j}}$) and scale-level errors (e_{satis}) can be considered types of simple measurement error, since they are uncorrelated with the errors and values of other constructs. Simple measurement error typical creates attenuation bias, if uncorrected for, causing estimates of association to be biased toward zero. Thus, simple measurement error is often considered to be of somewhat minimal concern, since it is not generally expected to produce false positives. There are important exceptions to this generalization, though, such as when a confounding variable has large simple measurement error—in such cases, failing to correct for the measurement error in the confounder could seriously bias main results in a variety of ways (including inflating estimates or creating false positives), since confounding effects will not be fully controlled for due to poor measurement (see Achen 1983; Freckleton 2011). Nonetheless, simple measurement error is generally considered to be a less serious problem for traditional regression than correlated error, or what I have called wicked errors.

Wicked measurement errors can cause serious problems for estimating associations, since they can easily lead to inflated estimates and false positives. In public administration, several studies in the past decade have brought growing awareness to the problem of common source variance—an issue that had been largely overlooked in this literature in prior decades (Meier & O'Toole 2012; George & Pandey 2017). Common source bias can arise for many reasons, but one simple example is social desirability bias. Because respondents vary in the degree to which their answers are affected by social desirability, and because many different constructs (including both independent and dependent variables) may be subject to social desirability bias, estimated associations for two unrelated constructs can easily yield positive results owing to common measurement error. Specifically, since responses to the measures of one construct can be a proxy for the susceptibility to social desirability bias by respondents, the variable can help to predict responses to another set of items that are also vulnerable to social desirability bias, even if these items reflect constructs that are unrelated to the original one.

One common justification for using a SEM framework is that it is good at dealing with measurement errors. However, I would argue that this statement is generally only true when considering Cronbach measurement errors. As a practical matter, it is often difficult to create plausible models in SEM for scale-level or wicked measurement errors—at least models that are identified and will converge. Furthermore, tools for handling simple measurement error exist (but are perhaps underutilized) in a traditional econometric framework in the form of errors-in-variables regression, so the SEM framework is not the only way to credibly account for simple measurement error. When it comes to more difficult-to-correct types of measurement error like scale-level or wicked errors, I would argue that the best solutions generally lie not in a particular statistical framework but instead in research design and careful interpretation of results or probing for robustness under alternative modeling assumptions.

SEM Versus the Traditional Econometric-Informed Approach to Attitudinal Research

Many researchers engage in research of attitudes using techniques other than the SEM framework. As a frame of reference, I think it will be useful to lay out what I have in mind as a typical process for studying attitudes through regression outside of SEM. I will refer to this process as the Traditional Framework. The traditional process I have in mind is as follows:

1. Calculate Cronbach's Alpha (or another reliability measure) for each multi-item scale
2. Run a factor analysis (exploratory or confirmatory) to check that for each scale, all items load strongly on a single dimension (or load distinctly on the expected number of dimensions)

3. For each scale, create a summative index (take the average of the items or add them all up), or create a factor index from a factor analysis
4. Enter these indices directly into a traditional regression model

Within an SEM framework, all of these steps may be combined into a single model estimation, or—maybe more commonly—broken down into at least two steps, starting with a confirmatory factor analysis (CFA) before creating a structural model.

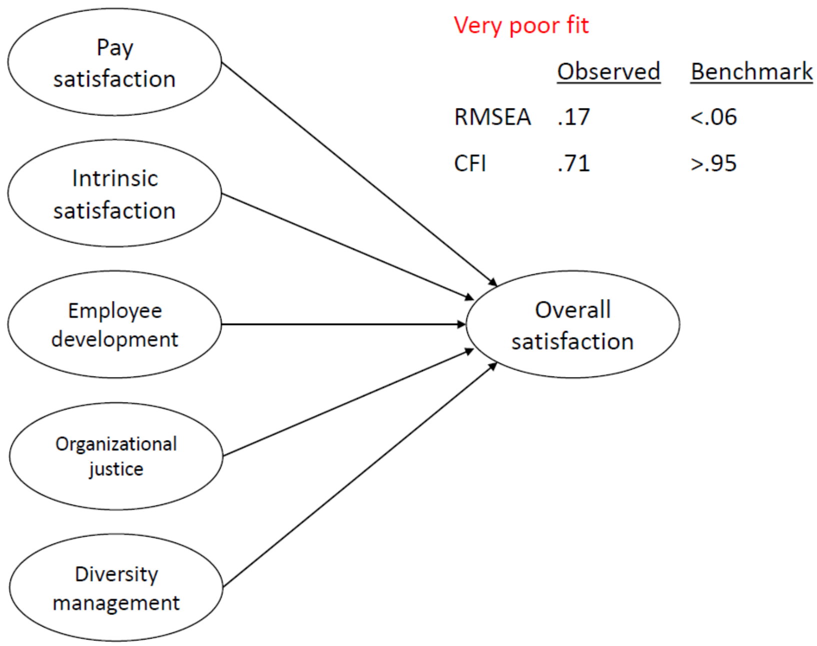
Introduction to SEM

For those unfamiliar with SEM, I offer a brief introduction to the framework. Essentially, SEM allows for combining a measurement model (roughly encompassing steps 1-3 in the Traditional Framework) with a structural model (step 4), such that they can be estimated together as one (large) single model. The structural model is similar to running a series of regressions, as is sometimes done in a path analysis using traditional regression models (e.g., OLS).

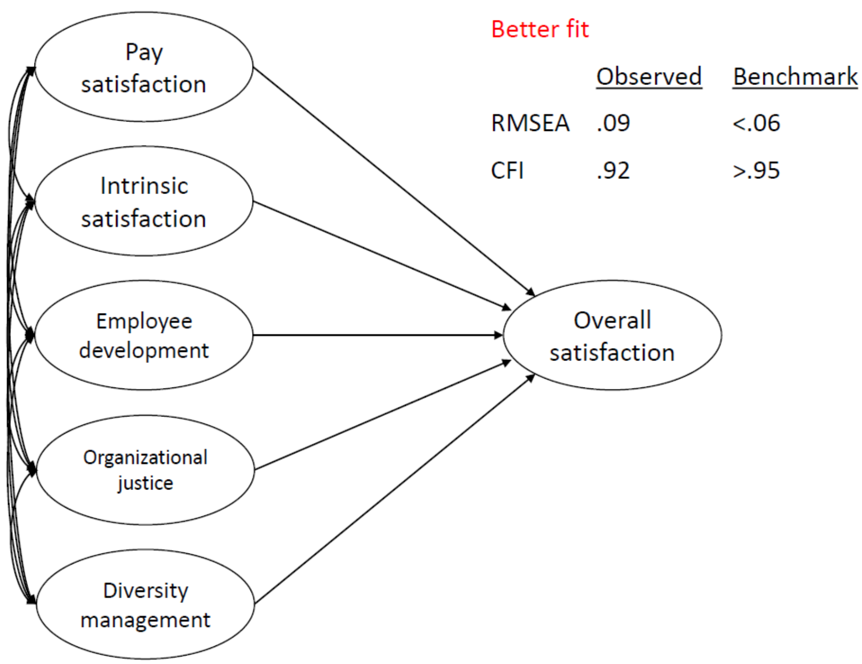
SEM has several similarities to traditional regression tools. For example, coefficients describing the associations between variables are interpreted in the same manner under either approach. One key difference, though, is that SEM is generally used to model a comprehensive system of relationships among several variables, spelling out all of the ways in which each variable is expected to be linked to every other variable. Contrast this with traditional regression, when one often focuses on one key dependent variable and examines how each independent variable might be associated with it, ignoring (to a large extent, anyway) the potential structure of how various independent variables might be linked to one another.

Figure 1. Example of fit indices from two alternative models

a. Model without linkages among independent variables



b. Model with linkages among independent variables



For those trained in an applied econometric tradition, perhaps one of the most unfamiliar aspects of the SEM approach is the use of fit indices. Fit indices stem directly from the focus on building a comprehensive system of relationships that is common within a SEM framework. Fit indices help to answer the question: in my model, did I omit any links that connect my variables to one another? We can think about links in terms of arrows since SEM models are often depicted visually with arrows connecting variables. Fit indices highlight the importance of noting which variables are *not* connected by arrows. The *lack* of an arrow connecting two variables indicates we are making the assumption that the two variables have a (direct) association of 0. There is an exception to this rule made for “exogenous” variables that we include in our model, which in practice often consists of a set of demographic characteristics that are assumed to be measured without any error. But at least for all attitudinal measures (usually treated as endogenous), unless we are willing to make the assumption that two variables are independent of one another, we generally need to draw an arrow between them under an SEM framework. To those trained in a (closed form) econometric tradition, this is a rather foreign idea since we are not used to having to make assumptions about the presence or absence of relationships among independent variables. When a model returns a poor fit, it means that the assumption that all variables not linked by arrows have no direct relationships is not supported by the data. More precisely, fit indices help to answer the question: can the links in my model explain all values in a full covariance matrix for all (endogenous) variables included in my analysis?

Figure 1 provides an illustration of how fit indices detect when there are missing links among variables. This illustration draws on data analyzed by Lee, Robertson, and Kim (2020), which I’ll discuss in more detail below under Example 1. But for now, the main thing to know about this example is that it is used to test how employee attitudes or perceptions about several aspects of their jobs are related to their overall job satisfaction. The two panes of Figure 1 depict two different SEMs. Only the structural parts of the model are shown in the figure, but we can assume for now that the measurement model (which describes how individual survey items are linked to the variables shown in the figure) fits the data reasonably well. In the top pane, we see that each independent variable is linked to the dependent variable of overall satisfaction, but no links have been created among the independent variables. This model yields fit indices that are considered to be very poor. In the bottom pane, we see that arrows have been added connecting all of the independent variables to one another. While these links are not of particular substantive interests given the research question, their inclusion dramatically improves the model fit.

While a fuller discussion of fit indices is beyond the scope of this paper, most introductions to SEM will provide detailed instructions on how to interpret fit indices and common thresholds for an acceptable fit. My emphasis here is on helping to provide an intuition for the higher-level question that SEM fit indices are seeking to help answer, since this is key to helping readers follow along with discussions of SEM applications. Beyond what I have already discussed, there are two further points worth briefly mentioning here regarding fit indices.

First, a good model fit does not mean that the arrows that have been drawn among variables are pointing in the correct (causal) direction. As we will see later in one of our examples, switching the direction in which an arrow points sometimes has no effect at all on the fit index. Within the SEM literature, this is a well-known instance of the broader issue that multiple models often yield equivalent model fit. As such, a good model fit does not indicate that the correct model has been specified—only that the specified model is a plausible one (and there may very well be several plausible alternatives).

Second, statistical programs can make suggestions about potential arrows to add to a model in order to improve the model fit. These suggestions come in the form of modification indices. Thus, the process of creating a SEM often consists of iterating through multiple models, perhaps making changes with the help of modification indices, until the researcher finds a model with satisfactory fit.

Guidelines for Analyzing Attitudes

In Table 1, I provide a brief overview of what I believe to be the main benefits and downsides of the SEM. My own opinion is that SEM is neither superior nor inferior to the traditional framework. Instead, the particularities of a given research project and its researchers will help to determine whether or not SEM is the preferred approach. The main benefits of SEM that I have identified relate to measurement error and mediation effects. At the same time, it is not so much that a Traditional Framework is unable to deal with these issues, as the workflow of SEM typically prods researchers to think about them with a level of care that is not always applied by researchers using the Traditional Framework. There are also some ways in which an SEM framework is genuinely more flexible and powerful when it comes to dealing with measurement error, but I suspect the added capabilities of SEM are often unlikely to be especially necessary or useful for most applied research applications.

Table 1. Should you use SEM?

Benefits of SEM	Downsides of SEM
• Forces you to think about measurement, error • When doing a horserace, can help you flag/confirm unique item-to-item common method variance issues • Encourages you to think about potential mediation and calculate total effects • Slightly more flexible approach to adjusting for measurement error than errors-in-variables, although still very limited (unclear to me how useful this really is)	• Requires certain types of data to work well (e.g., clean scales with at least 3 items per variable) • Complexity increases exponentially as variable list grows • Barriers to entry with technical trouble-shooting • Not very flexible (e.g., can be hard to deal with an item that behaves badly psychometrically that you decide to keep anyway) • Easily misunderstood as testing for causal direction

My broader advice on how to analyze attitudes is outlined as a set of guidelines in Table 2. These guidelines apply regardless of which framework of analysis is being used. My hope is that these guidelines will encourage researchers to engage more deeply with fundamental research design issues that should be confronted regardless of the estimating framework employed. Rather than discussing these guidelines in detail here, I will attempt to demonstrate some of the ways they can be applied through some examples in the following sections.

Table 2. Guidelines for analyzing attitudes

Steps to think through
1. Choose an analytical approach • Do you want to estimate independent associations (potential causal paths), or are bivariate associations more appropriate? • **Try to use tools you understand well** • If you have good multi-item scales for all (most) attitudinal measures, SEM may be a good option • If using SEM, you can always check if you get similar results running a similar (simpler) set of parallel analyses with traditional regression; just be sure to standardize all coefficients for comparability

2. Check for clear, distinct variables
• Are constructs clearly and distinctly defined?
• Do the measures of variables have good discriminant and convergent validity?
    o Examine item wording qualitatively to assess validity
    o Can run exploratory or confirmatory factor analysis (or just examine a correlation matrix)
    o Can calculate a Cronbach’s alpha, although maximizing this is not always desirable (e.g., items overly repetitive, don’t cover construct’s full scope)
• Imperfect measures can be okay, so long as you can explain/defend what is being measured

3. Don’t trust too much in statistical significance • Highly subjective survey items often correlate with one another to some extent, even if there is not a very meaningful relationship among the underlying variables, so finding a significant association is not necessarily meaningful • Real standard errors are probably larger than what your estimates show (especially if using OLS)
4. Focus on relative strength of associations, if appropriate

- **Set up a horserace** by having several variables from the same survey, and seeing which ones have the strongest relationships with the dependent variable
- If we assume all attitudinal variables are equally subject to a constant common method variance factor, a horserace among these variables should be valid (even if some variables produce false positives on significance tests)

5. Be aware of factors that might distort a horserace

- Associations tend to be artificially stronger:
  - Across items with common (or similar) wording (in prompt or response options)
  - Across items that appear nearby each other on a survey
  - When variables more precisely measured (such as indices with more items)
  - See other factors in Table 2 of Podsakoff et al. (2003)
- Account for these confounding factors qualitatively when interpreting results, or use SEM (or errors-in-variables regression) to try to correct for them

6. Be sure to consider total effects

- With bivariate correlations, you should already get a sense of the total magnitude of association
- With multiple regression (including within SEM), **consider possible mediating relationships**
  - Avoid controlling for potential mediators (except in the context of a mediation analysis where you calculate indirect effects)
  - When uncertain, see how results change under different assumptions about causal orderings
  - Think carefully about what it would mean to see a shock in one independent variable while holding others constant

Example 1: Employee satisfaction (Lee, Robertson, & Kim 2020)

Lee, Robertson, and Kim's (2020) study of employee satisfaction in the U.S. federal government provides a useful example of a study of attitudes. Their study draws on a large survey of federal employees and makes use of techniques from a traditional econometric tradition. In attempting to reanalyze their data under an SEM framework, I immediately encountered the problem that the data was not particularly well-suited to an SEM analysis because of insufficient items for several multi-item constructs (e.g., variables with only two or three indicators). As such, I encountered convergence difficulties, where for many SEMs I initially attempted to estimate, the software could not find a solution. This result in and of itself is useful in terms of illustrating some of the differences between SEM and the Traditional Framework, with the latter being able to accommodate multi-item measures with very few indicators. But to demonstrate other ways in which the two approaches differ/overlap, I decided to focus in this paper on reanalyzing a subset of the attitudinal independent variables from Lee, Robertson, and Kim's (2020) study, in order to allow for a comparison of results under SEM versus the Traditional Framework. The measures I examine here are shown in Table 3.

Table 3. Measure of attitudinal variables

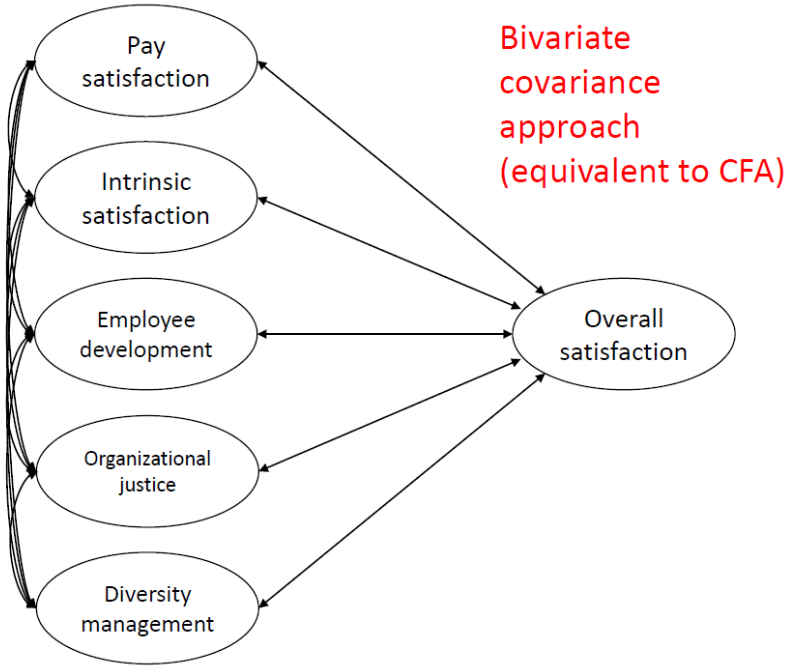
Variable	Questions	α
Overall satisfaction	Considering everything, how satisfied are you with your job?	.908
	Considering everything, how satisfied are you with your organization?	
	I recommend my organization as a good place to work.	
Intrinsic satisfaction	My work gives me a feeling of personal accomplishment.	.823
	I like the kind of work I do.	
	My talents are used well in the workplace.	
	The work I do is important.	
Employee development	I am given a real opportunity to improve my skills in my organizations.	.862
	My supervisor provides me with opportunities to demonstrate my leadership skills.	
	Supervisors in my work unit support employee development.	
	How satisfied are you with the training you receive for your present job?	
Organizational justice	I can disclose a suspected violation of any law, rule, or regulation without fear of reprisal.	.854
	Arbitrary action, personal favoritism, and coercion for partisan political purposes are not tolerated.	
	Prohibited personnel practices (e.g., illegally discriminating for or against any employee/applicant, obstructing a person's right to compete for employment, knowingly violating veterans' preference requirements) are not tolerated.	
Pay satisfaction	Considering everything, how satisfied are you with your pay?	NA
Diversity management	Policies and programs promote diversity in the workplace (e.g., recruiting minorities and women, training in awareness of diversity issues, mentoring).	.791
	My supervisor is committed to a workforce representative of all segments of society.	
	Supervisors work well with employees of different backgrounds.	

Note: Adapted from Table 1 of Lee, Robertson, & Kim (2020)

Overall job satisfaction can be considered the dependent variable, and the other measures are all potential antecedents of satisfaction, although it is also quite possible that there is reverse or reciprocal causality. Given the observational cross-sectional data, the absence of clear *a priori* reasoning that can justify an assumption about causal direction, and even potential conceptual overlap among some of the variables, this strikes me as a case where it may make sense to approach the analysis as a largely descriptive exercise, trying to map out which variables are most closely linked in a bivariate sense but not attempting to disentangle independent associations. Figure 2 demonstrates the distinction between these two approaches, depicted with arrows as we often use to graphically illustrate models within an SEM framework.

Figure 2. Bivariate associations versus independent associations

a. Bivariate associations



b. Independent associations

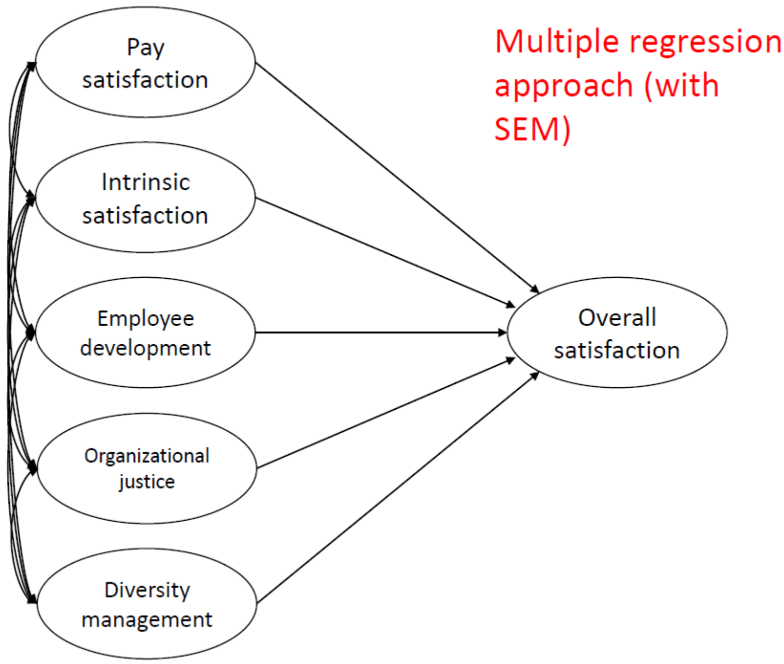


Table 4 shows the results of estimating such models, both within an SEM framework and within the Traditional Framework. The independent variables are listed in order of their estimated reliability (based on Cronbach's alpha), with pay satisfaction assumed to have a perfect reliability of 1 (since it is a single-item measure and thus a reliability coefficient cannot be estimated). Since the SEM models will account for Cronbach errors, and since random errors (including Cronbach errors) will generally lead to attenuation bias, we expect that associations will generally look a bit stronger after correcting for Cronbach errors. Variables measured with larger Cronbach errors should get a bigger correction.

Table 4. Comparing traditional approaches (Pearson correlations and OLS) with SEM

DV: Overall satisfaction		Bivariate associations (covariances)			Independent associations (multiple regression)		
	Cronbach's α	(1) Corr.	(2) SEM (CFA)	Increase w/ SEM ((2)-(1))	(3) OLS	(4) SEM	Increase w/ SEM ((4)-(3))
Pay satisfaction	NA	0.48	0.51	0.03	0.14	0.13	-0.01
Employee development	0.86	0.75	0.82	0.07	0.24	0.19	-0.05
Organizational justice	0.85	0.70	0.77	0.07	0.21	0.24	0.03
Intrinsic satisfaction	0.82	0.74	0.84	0.10	0.36	0.46	0.10
Diversity management	0.79	0.68	0.8	0.12	0.09	0.03	-0.06

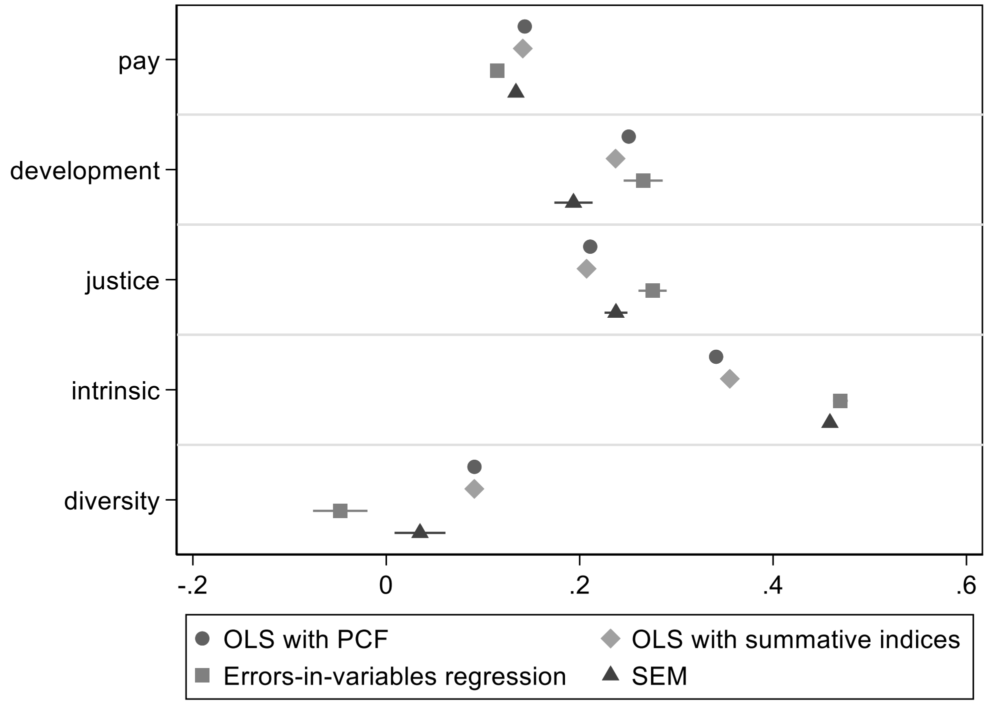
Note: All coefficients/covariances are standardized. Columns (1) and (3) use summative indices.

Starting from the estimates of bivariate associations, we can see that the simple correlation coefficients generally provide very similar estimates to the SEM-generated estimates of bivariate associations, although the SEM estimates tend to be a bit larger. The column showing the differences between the two sets of estimates ((2)-(1)) makes it clear that variables with larger Cronbach errors (toward the bottom of the table) see a larger increase in estimated size of association in the SEM.

The results for independent associations are a bit messier. While the OLS results are still not far off from the SEM estimates, in several cases the associations are estimated to be weaker in the SEM models that account for Cronbach error. This owes in part to the fact that the effects of random measurement error are much more difficult to anticipate in a multiple regression context, especially when there is substantial multicollinearity (Achen 1983; Freckleton 2011).

I also tried estimating independent associations in two other ways: using OLS but relying on factor indices (rather than summative indices) from a principal components factor (PCF) analysis, as well as an errors-in-variables regression in which I used Cronbach's alpha values to estimate the reliability of each variable measure. As Figure 3 makes clear, the errors-in-variables model (part of the traditional econometrics tradition) yields adjustments for measurement errors which are fairly different from the adjustments made by the SEM in this case. Nonetheless, both models that make corrections for Cronbach errors indicate significantly wider confidence intervals than the OLS models for several coefficients, suggesting that OLS dramatically overstates the precision of estimates for these variables by ignoring measurement error.

Figure 3. Comparison of estimates from alternative models of overall satisfaction



Finally, I demonstrate the comparative flexibility of SEM in addressing measurement errors by adding a new latent variable reflecting common method variance associated with inclusion of the word "satisfied" in the item prompt. As can be seen from item wordings listed in Table 3, four item prompts contain the word "satisfied": two items measuring overall satisfaction, one measuring recognition for good work, and one measuring employee development. Figure 4 provides a visual depiction of this new SEM model.

Figure 4. SEM model with latent factor for common method variance

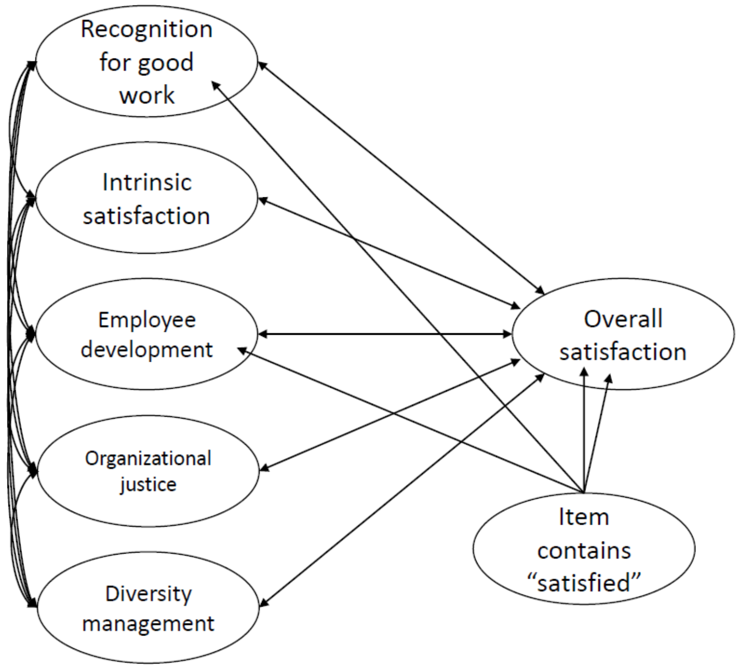


Table 5 contains the results of this model. As can be seen, results barely change. However, the Recognition for Good Work gets a slightly smaller association, suggesting that the common method variation led to a slightly inflated association before. The slight bumps in size of estimated association for Organizational Justice and Diversity Management suggest that these associations were (very slightly) underestimated because they lacked the common source variation found in two other predictors of Overall Satisfaction.

Table 4. Comparing SEM (bivariate) association estimates with and without a latent factor for common method variance

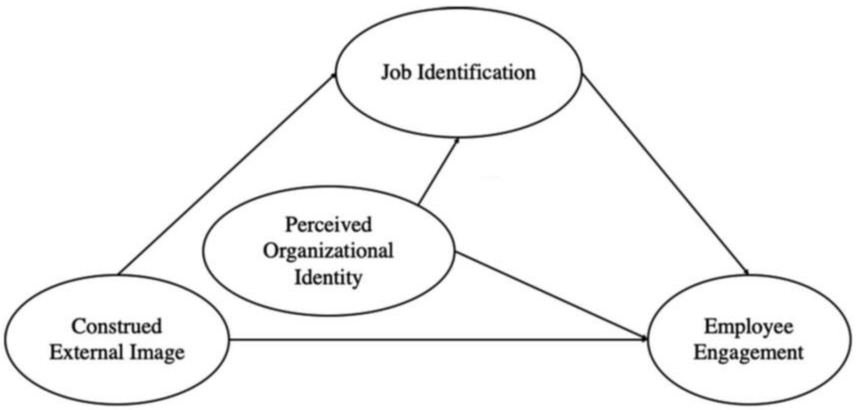
	(1) SEM (CFA)	(2) SEM w/ “satisfied” var.	Increase w/ “satisfied” var. ((2) – (1))
Employee development*	0.82	0.82	0.00
Organizational justice	0.77	0.79	0.02
Recognition for good work*	0.86	0.85	-0.01
Intrinsic satisfaction	0.84	0.84	0.00
Diversity management	0.81	0.82	0.01

Note: * indicates variable contains an item with the word “satisfied.” All coefficients/covariances are standardized.

Example 2: Employee engagement (Hameduddin & Lee 2021)

Hameduddin and Lee (2021) provide us with another useful example. Their model of employee engagement (based on another survey of U.S. federal employees: the U.S. Merit Systems Protection Board's 2016 Merit Principles Survey) is depicted in Figure 5. The items measuring their variables are shown in Table 5.

Figure 5. A model of employee engagement



Note: Adapted from Figure 2 of Hameduddin & Lee (2021)

Table 5. Measures of employee engagement and its antecedents

Employee Engagement (ee) 1. 'My work gets me energized and excited' 2. 'I put my full physical energy into doing my work tasks' 3. 'It is easy for me to become happily immersed in my work' 4. 'I feel engaged in my job'
Job Identification (ji) 1. 'My work gives me a good opportunity to do things I am passionate about' 2. 'My work supports a purpose, cause, or mission that is important to me' 3. 'My work is consistent with my core values and beliefs' 4. 'My work is consistent with my personal sense of purpose or calling'
Perceived Organizational Identity (poi) 1. 'My agency is successful at accomplishing its mission' 2. 'I would recommend my agency as a place to work'
Construed External Image 1. 'Public support for your organization's mission and work' 2. 'Public perception of your organization's performance'

A simple confirmatory factor analysis based on the variables in Table 5 indicates that fit is slightly below expected thresholds (RMSE=.07, CFI=.96). The biggest value for a modification index is for linking item ee4 to Organizational Identity. This happens to align largely with a problem we might flag if analyzing these same items with exploratory factor analysis. Table 6 shows that the item ee4 does not load well with the other ee items and instead has the strongest loading with the poi items. It turns out that the first three ee items (ee1-ee3) were all asked on the same section of the survey, while ee4 was asked separately in a section of the survey that also appeared next to items that asked about attitudes toward more global aspects of the organization, similar to the poi items. Thus, both PCA and exploratory factor analysis are picking up on a research design issue with the survey and perhaps a lack of strong discriminant validity among variables (arguably reflecting a deeper lack of consistent and clear conceptual distinctions among constructs).

Table 6. Exploratory factor analysis (rotated PCA)

Variable	Factor1	Factor2	Factor3	Factor4	Uniqueness
ee1	0.4283	0.6711	0.3223	0.1075	0.2508
ee2	0.1948	0.8093	0.0099	0.1006	0.2969
ee3	0.3397	0.7375	0.2758	0.1045	0.2537
ee4	0.4046	0.4208	0.5670	0.0691	0.3330
cei1	0.1006	0.0792	0.1324	0.9385	0.0853
cei2	0.0665	0.0744	0.1636	0.9362	0.0868
poi1	0.1991	0.1004	0.7808	0.2946	0.2538
poi2	0.2562	0.1670	0.8149	0.1946	0.2045
ji1	0.7347	0.3527	0.2287	0.1018	0.2731
ji2	0.8001	0.1849	0.1767	0.0924	0.2858
ji3	0.8030	0.1917	0.2396	0.0988	0.2513
ji4	0.8242	0.2610	0.1799	0.0954	0.2110

I again tried comparing four different methods of estimating independent associations of the dependent variable with key constructs, as well as with a set of control variables (demographic factors). The results are shown in Figure 6. Across models, results are quite similar, although the errors-in-variables approach more closely mirrors SEM than the OLS models.

Figure 6. Comparison of estimates from alternative models of employee engagement and job identification

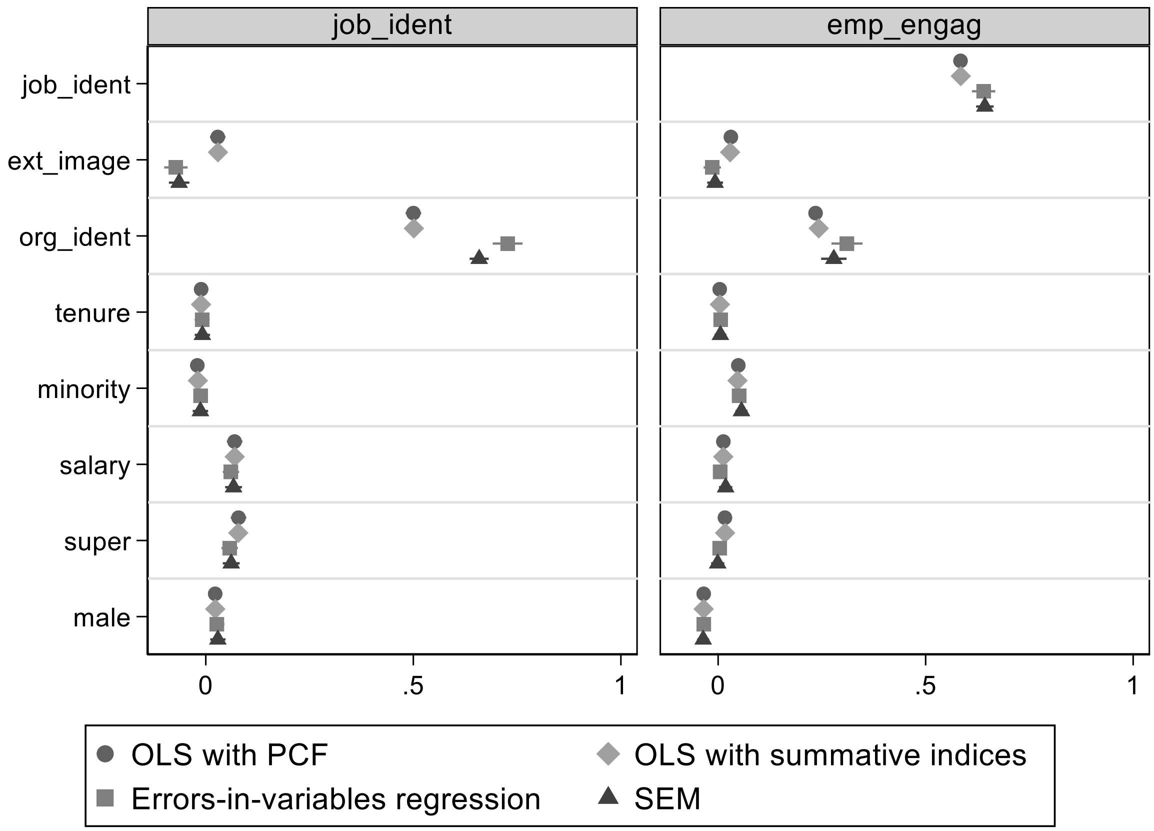
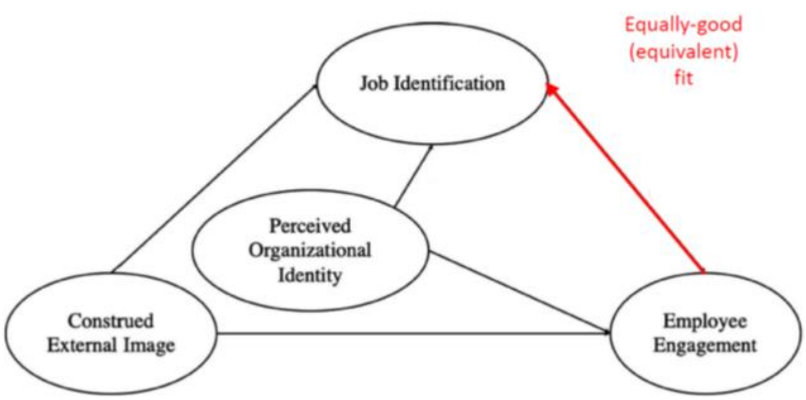


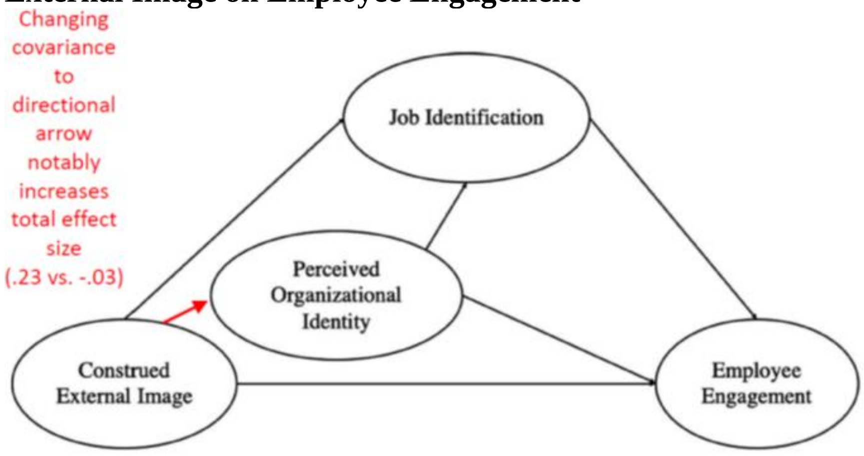
Figure 7. Alternative model with equally good fit



This dataset can also be used to demonstrate the issue of equivalently-plausible models. Figure 7 shows an alternative to the base model (Figure 5)—an alternative which generates equivalent model fit, despite making a different assumption about direction of causality.

Finally, I wish to highlight the importance of carefully considering decisions about causal directions assumed for associations in SEM by comparing a model with a covariance link (nondirectional arrow) between Construed External Image and Perceived Organizational Identity (as in Figure 5, though the covariance arrow was not shown) versus a model with a directional error, as shown in Figure 8. Inclusion of this causal connection leads to finding a dramatically larger effect of construed external image on employee engagement (.23, versus -.03 in the original model).

Figure 8. Effects of adding a directional error on the estimated total effect of Construed External Image on Employee Engagement



References

Achen, Christopher H. 1983. "Toward Theories of Data: The State of Political Methodology." In *Political Science: The State of the Discipline*, ed. Ada W. Finifter. Washington, DC: American Political Science Association, 69-93.

Freckleton, Robert P. 2011. "Dealing with collinearity in behavioural and ecological data: model averaging and the problems of measurement error." *Behavioral Ecology and Sociobiology* 65: 91-101.

George, Bert, and Sanjay K. Pandey. 2017. "We know the Yin—But where is the Yang? Toward a balanced approach on common source bias in public administration scholarship." *Review of public personnel administration* 37, no. 2: 245-270.

Grant, Adam M. 2008. "Does intrinsic motivation fuel the prosocial fire? Motivational synergy in predicting persistence, performance, and productivity." *Journal of applied psychology* 93, no. 1: 48-58.

Hameduddin, Taha, and Shinwoo Lee. 2021. "Employee engagement among public employees: Examining the role of organizational images." Public Management Review 23, no. 3: 422-446.

Lee, Hyung-Woo, Peter J. Robertson, and Kitai Kim. 2020. "Determinants of job satisfaction among US federal employees: An investigation of racial and gender differences." Public Personnel Management 49, no. 3: 336-366.

Marvel, John D. 2016. "Unconscious bias in citizens' evaluations of public sector performance." *Journal of Public Administration Research and Theory* 26, no. 1: 143-158.

Meier, Kenneth J., and Laurence J. O'toole. 2012. "Subjective organizational performance and measurement error: Common source bias and spurious relationships." *Journal of Public Administration Research and Theory* 23, no. 2: 429-456.

Podsakoff, Philip M., Scott B. MacKenzie, Jeong-Yeon Lee, and Nathan P. Podsakoff. 2003. "Common method biases in behavioral research: a critical review of the literature and recommended remedies." Journal of applied psychology 88, no. 5: 879.