Psychological Methods
2014, Vol. 19, No. 2, 300–315

1082-989X/14/$12.00 DOI: 10.1037/a0033805

Manifest Variable Path Analysis: Potentially Serious and Misleading Consequences Due to Uncorrected Measurement Error

David A. Cole and Kristopher J. Preacher
Vanderbilt University

Despite clear evidence that manifest variable path analysis requires highly reliable measures, path analyses with fallible measures are commonplace even in premier journals. Using fallible measures in path analysis can cause several serious problems: (a) As measurement error pervades a given data set, many path coefficients may be either over- or underestimated. (b) Extensive measurement error diminishes power and can prevent invalid models from being rejected. (c) Even a little measurement error can cause valid models to appear invalid. (d) Differential measurement error in various parts of a model can change the substantive conclusions that derive from path analysis. (e) All of these problems become increasingly serious and intractable as models become more complex. Methods to prevent and correct these problems are reviewed. The conclusion is that researchers should use more reliable measures (or correct for measurement error in the measures they do use), obtain multiple measures for use in latent variable modeling, and test simpler models containing fewer variables.

Keywords: path analysis, measurement error, reliability, goodness-of-fit, parameter estimation

Using path analysis to model relations among manifest (not latent) variables carries the assumption that the measures are reliable manifestations of the constructs they represent (Bollen, 1989). Methodologists have described a variety of problems that can emerge when this assumption is not met (e.g., Bollen, 1989; James, Mulaik, & Brett, 1982; Kenny, 1979; Ledgerwood & Shrout, 2011; Rigdon, 1994; Rubio & Gillespie, 1995; Wansbeek & Meijer, 2000). Nevertheless, our review of the literature reveals that the publication of path analyses with fallible variables remains quite common, even in premier journal outlets (we elaborate on this point in the following section). In the current article, we demonstrate that even modest amounts of measurement error[1] can lead to substantial overestimation of some path coefficients, substantial underestimation of other path coefficients, and significant evidence of model misspecification even when the model perfectly represents the true relations among the underlying constructs.[2] We demonstrate that these problems exist in very simple models but become more serious, more numerous, and less tractable as models become more complex.

This article has six parts. First, we conduct a brief literature review showing the prevalence of manifest variable path analysis across our discipline. Second, using a concept we call the “fallible triangle,” we provide an intuitive explanation of the reasons why measurement error affects manifest variable path analyses so profoundly. Third, we develop a general method for understanding over- and underestimation in complex models. Fourth, we examine the effect of measurement error on goodness-of-fit. Fifth, we highlight these points with an example. Finally, we review some of the methods available for coping with such problems.

The Prevalence of Path Analyses: Literature Review

Path analyses with fallible variables are relatively common in the behavioral sciences. To document the extent to which path analyses are used across a wide range of psychological subdisciplines, we reviewed the recent issues of seven major American Psychological Association journals: *Developmental Psychology*, *Health Psychology*, *Journal of Abnormal Psychology*, *Journal of Applied Psychology*, *Journal of Educational Psychology*, *Journal of Family Psychology*, and *Journal of Personality and Social Psychology*. Across 44 issues published in 2011, we found 91 articles that reported results of at least one path analysis; that is, 11.7% of the publications in these journals included path analysis, for an average of more than two manifest variable path analytic articles per issue (see Table 1). In this review, we excluded articles that used multiple regression analyses that were not embedded in a larger path analytic framework (even though multiple regression is itself a special case of path analysis); consequently, our numbers underestimate the actual prevalence of path analysis with fallible measures.

This article was published Online First September 30, 2013.

David A. Cole and Kristopher J. Preacher, Department of Psychology and Human Development, Vanderbilt University.

This research was supported by a gift from Patricia and Rodes Hart to David A. Cole.

Correspondence concerning this article should be addressed to David A. Cole, Department of Psychology and Human Development, Vanderbilt University, Nashville, TN 37203-5721. E-mail: david.cole@vanderbilt.edu

[1] Throughout this article, we use the term *measurement error* to refer to random and normal measurement error, not systematic or nonnormal measurement error (which may have different implications that are also worthy of study).

[2] Throughout, we assume that all path coefficients are non-negative. In this context, the words *attenuation* and *underestimation* refer to the shrinkage toward zero of path coefficient estimates due to measurement error. The words *inflation* and *overestimation* refer to the expansion away from zero of path coefficient estimates.

Table 1
Papers Using Manifest Variable Path Analyses, Published in Seven 2011 Journals

Journal	No. of journal issues reviewed	No. of articles reviewed	No. of articles with path analyses	Average no. of manifest variables per model	Range of manifest variables per model
Developmental Psychology	6	155	10	6.75	2–13
Health Psychology	6	93	11	4.43	2–16
Journal of Abnormal Psychology	4	93	9	7.17	2–18
Journal of Applied Psychology	6	93	14	6.04	3–11
Journal of Educational Psychology	4	63	7	7.60	3–18
Journal of Family Psychology	6	108	14	7.95	2–14
Journal of Personality and Social Psychology	12	174	26	5.51	3–13
Total	44	779	91		

Calculating the effects of measurement error in these articles is difficult because approximately 64% of the articles were missing reliability information for at least one of the modeled variables. Of the 535 modeled variables for which reliabilities were reported, the median reliability was 0.84; however, 33.6% had reliabilities less than 0.80, and 8.3% had reliabilities less than 0.70. Although many of these values may not seem particularly low, our later analyses show that reliabilities as large as 0.80 and even 0.90 can be cause for concern. In no case did authors attempt to utilize any procedure to correct for measurement error (Bollen, 1989; Coffman & MacCallum, 2005; Kishton & Widaman, 1994; Little, Cunningham, Shahar, & Widaman, 2002; Stephenson & Holbert, 2003). In approximately 13% of the articles, a path analysis contained a mix of manifest and latent variables, creating the potential for large reliability differences in various parts of the model. (Models in which all manifest variables loaded onto latent variables were not included in this review.) The number of manifest variables per model was highly variable, ranging from as few as two or three to as many as 16 or 18 in some journals (see Table 1). The overall mean was 6.1 manifest variables per model ($SD = 3.4$). In general, articles that had more variables in their most complex model included measures with lower reliabilities ($r = -.31$, $p < .005$, $n = 91$ studies). In a nutshell, manifest variable path analyses are very common and typically are conducted without correction for measurement error (and often without all of the necessary information to conduct such corrections after the fact).

Why Measurement Error Affects Path Analyses

In this section, we explain how measurement error can simultaneously inflate and attenuate estimates of certain path coefficients in manifest variable path analyses and why this problem gets worse as models become more complex. In this section and throughout the article, we focus primarily on standardized path coefficients. We made this decision in large part because virtually all of the published applications of path analysis that we reviewed focused on the interpretation of standardized path coefficients, often to the complete exclusion of unstandardized coefficients. Although much of what we say in this article generalizes to the unstandardized case, some things do not (see Blalock, 1964; Duncan, 1975; Kenny, 1979). Broader arguments about the advantages and disadvantages of focusing on standardized versus unstandardized coefficients have been well articulated elsewhere (Baguley, 2009; Greenland, Schlesselman, & Criqui, 1986; Kim & Ferree, 1981; G. King, 1986). Our choice to concentrate primarily on the standardized case is based upon current practices and should not be taken as an endorsement of one method over the other.

The Fallible Triangle Problem

Measurement error can lead to the spurious underestimation of some paths and to the spurious overestimation of others.[3] To show this, let us begin with Model 1a, in which F is a function of D and E, and E is a function of D (as shown in Figure 1). If D, E, and F were measured perfectly (i.e., if they were error-free latent variables), their correlations would be $r_{DE} = 0.40$, $r_{DF} = 0.40$, and $r_{EF} = 0.58$. From these, the standardized path coefficients between the latent variables can be calculated:

$$\beta_{ED} = r_{DE} = 0.40, \tag{1}$$

$$\beta_{FE \cdot D} = (r_{EF} - r_{DF} r_{DE}) / (1 - r^2_{DE}) = 0.50, \text{ and} \tag{2}$$

$$\beta_{FD \cdot E} = (r_{DF} - r_{EF} r_{DE}) / (1 - r^2_{DE}) = 0.20. \tag{3}$$

For the purposes of this article, we refer to the relations among D, E, and F as a triangle (for obvious reasons). Actually, this is an "infallible" triangle because D, E, and F are measured without error (i.e., D, E, and F may be considered latent variables).

If any or all of these three variables were not perfectly measured, then we would refer to the model as a *fallible* triangle. In fallible triangles with at least one dependent variable, measurement error in one of the independent variables can spuriously inflate the path connecting the other two variables. This fact is well documented, albeit rarely mentioned in empirical articles involving such analyses (Hoyle & Kenny, 1999; Kenny, 1979; Rigdon, 1994; Wansbeek & Meijer, 2000; Wolfle, 1979, 1980). A classic example of path coefficient inflation occurs in mediation models (Hoyle & Kenny, 1999; Ledgerwood & Shrout, 2011). To see this, let us imagine that any one of the three latent variables (D, E, or F) might be represented by a fallible version of itself (D′, E′, or F′, respectively). Let us further imagine that each fallible measure correlated 0.7 with (or loaded 0.7 onto) its corresponding latent variable. When the mediator is measured with error but the other

[3] Throughout this article (unless explicitly stated otherwise), we assume unidimensional, reflective indicators (i.e., each indicator taps only one underlying construct) such that the reliability of each measure represents its proportion of true-score variance.

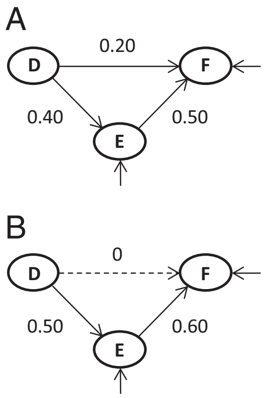

Figure 1. A. Model 1a: Path diagram with standardized path coefficients that assume D, E, and F are measured without error. B. Model 1b: Path diagram where E completely explains the correlation between D and F.

two variables (D and F) are not, the coefficient representing the direct effect of D→F is inflated. For example in our Model 1a, using only E′ to represent E will attenuate two out of three correlations, $r_{DE'}$ and $r_{E'F}$, such that $r_{DE'} = r_{DE}\sqrt{r_{E'E'}} = (0.40)(0.70) = 0.28$ and $r_{E'F} = r_{EF}\sqrt{r_{E'E'}} = (0.58)(0.70) = 0.406$. When $r_{DE'}$ and $r_{E'F}$ replace r_{DE} and r_{EF} in Equation 3, $\beta_{FD \cdot E'}$ is estimated to be $(0.40 - 0.406 * 0.28) / (1 - 0.28^2)$, or 0.311 instead of 0.20. In analogous fashion, using only D′ to measure D will inflate the E→F path estimate from 0.50 to 0.54. Depending upon the sample size, increasing measurement error could inflate estimates from nonsignificant to statistically significant levels.

More often acknowledged in empirical articles is the fact that estimates of correlations and path coefficients can also be attenuated by measurement error (e.g., Blalock, 1965; Hoyle & Kenny, 1999; Ledgerwood & Shrout, 2011). In our Model 1a, if D were measured only by D′ (but E and F remained error-free), both $r_{D'F}$ and $r_{D'E}$ would underestimate r_{DF} and r_{DE}, respectively (i.e., both correlations would be 0.28 not 0.40). Consequently, the D→E path estimate would drop from $\beta_{ED} = 0.40$ to $\beta_{ED'} = 0.28$, and the D→F path estimate would drop from $\beta_{FD \cdot E} = 0.20$ to $\beta_{FD' \cdot E} = (0.28 - 0.58 * 0.28) / (1 - 0.28^2) = 0.13$. Alternatively, if E were measured only by E′ (but D and F were error-free), r_{DE} and r_{EF} would be underestimated by $r_{DE'} = 0.40 * 0.70 = 0.28$ and $r_{E'F} = 0.58 * 0.70 = 0.406$, respectively, which would then attenuate the D→E path from $\beta_{ED} = 0.40$ to $\beta_{E'D} = 0.40 * 0.70 = 0.28$ and the E→F path from $\beta_{FE} = 0.50$ to $\beta_{FE' \cdot D} = (0.406 - 0.40 * 0.28) / (1 - 0.28^2) = 0.319$. Finally, if F were measured by only F′ (but D and E were error-free), then $r_{EF} = 0.58$ would drop to $r_{EF'} = 0.58 * 0.70 = 0.406$ and $r_{DF} = 0.40$ would drop to $r_{DF'} = 0.40 * 0.70 = 0.28$, which would then attenuate the E→F path from $\beta_{FE \cdot D} = 0.50$ to $\beta_{F'E \cdot D} = 0.50 (0.406 - 0.28 * 0.40) / (1 - 0.40^2) = 0.35$ and the D→F path from $\beta_{FD \cdot E} = 0.20$ to $\beta_{F'D \cdot E} = 0.20(0.28 - 0.406 * 0.40) / (1 - 0.40^2) = 0.14$. Depending upon sample size, increasing measurement error could attenuate some of these estimates from significant to nonsignificant levels.

We can think of unreliability in any *single* variable as applying "pressure" on the estimation method to inflate or attenuate certain path coefficients in the model. When *multiple* variables in the model are fallible, *multiple* inflationary and/or attenuating pressures can simultaneously affect some path coefficients. Graphically, we depict each of these pressures with small upward and downward arrows adjacent to each of the paths in Model 2 (see Figure 2). The arrows are subscripted to indicate which pressure is driven by unreliability in which variable. (Similar things happen in models where fallible indicators of D and E are only correlated and not connected by a directional path.)

Implicit Fallible Triangles

Fixing certain path coefficients to zero does not eliminate fallible triangles; it just makes their effects more insidious. We call triangles in which one or more paths are fixed to zero "implicit fallible triangles" (to distinguish them from "explicit triangles" in which all three paths are free). Although fixing a path to zero (obviously) eliminates the possibility of its being overestimated, it does not eliminate the pressure to do so. Let us imagine Model 1b in Figure 1 in which the mediator E completely explains the relation between D and F when all variables are latent. For example, $r_{DF} = \beta_{ED}\beta_{FE} = 0.50 * 0.60 = 0.30$ and $\beta_{FD} = 0$. Anticipating this, the investigator fixes the D→F path to zero. As shown above, measuring the mediator with a single fallible measure (E′) will attenuate estimates of the standardized path coefficients that represent the indirect effect of D on F through E ($\beta_{E'D} = 0.35$ and $\beta_{FE' \cdot D} = \frac{0.42 - (0.35)(0.30)}{1 - 0.35^2} = 0.39886$), so that their product $(0.35 * 0.39886 = 0.1396)$ will no longer explain the 0.30 correlation between latent D and latent F. The difference between 0.1396 and 0.30 cannot be resolved by inflating the direct effect of D on F, as that path is fixed to zero. Instead, the discrepancy generates evidence of specification error and model misfit. Post hoc diagnostics must point to the fixed D→F path as the source of the problem, as it is the only constrained path. If the investigator were to relax this constraint, the path coefficient would then be inflated relative to its original (and proper) value of zero. Consequently, the investigator will end up with either a free but inflated path coefficient or a model that fits poorly. Under some circumstances, estimates of truly zero path coefficients can be inflated sufficiently to become statistically significant.

Generalizations About Complex Models

In this section, we first describe how the pressures to inflate or attenuate paths proliferate when models contain more variables

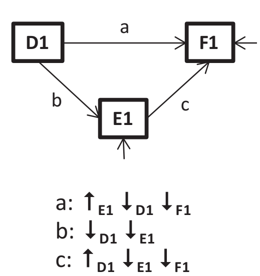

Figure 2. Model 2: An explicit fallible triangle, where the up and down arrows signify the susceptibility of standardized path coefficients to over- and underestimation, respectively, due to measurement error in the variable designated in the subscript.

and more paths. Second, we present a collection of principles that highlight how problems of parameter over- and underestimation become more numerous and more severe as models become increasingly complex.

Model Complexity Makes Matters Worse

In this section, we show that the pressures exerted by measurement error to over- and/or underestimate path coefficients can become more numerous, more serious, and less tractable as models become more complex. This happens in part because the paths in more complex models are likely to be part of a larger number of fallible triangles, each of which represents a vehicle for its over- and/or underestimation. In Figure 3, consider several increasingly complex examples, culled from recently published articles. Model 3a has four variables with four fallible triangles (three of which are implicit).[4] Model 3b has five variables with 10 fallible triangles (six of which are implicit). Model 3c represents a six-variable model with 20 fallible triangles (19 of which are implicit). The maximum number of fallible triangles among p variables can be calculated as the number of ways p things can be grouped r-at-a time, where $r = 3$ (for the three variables in a triangle), using the formula for combinations: $C(p,r) = p! / (r!(p - r)!) = p! / (3!(p - 3)!)$, in general. Examining the small over- and underestimation arrows in Figure 3 reveals that increasing the complexity of a model renders more paths susceptible to more inflation and attenuation effects.

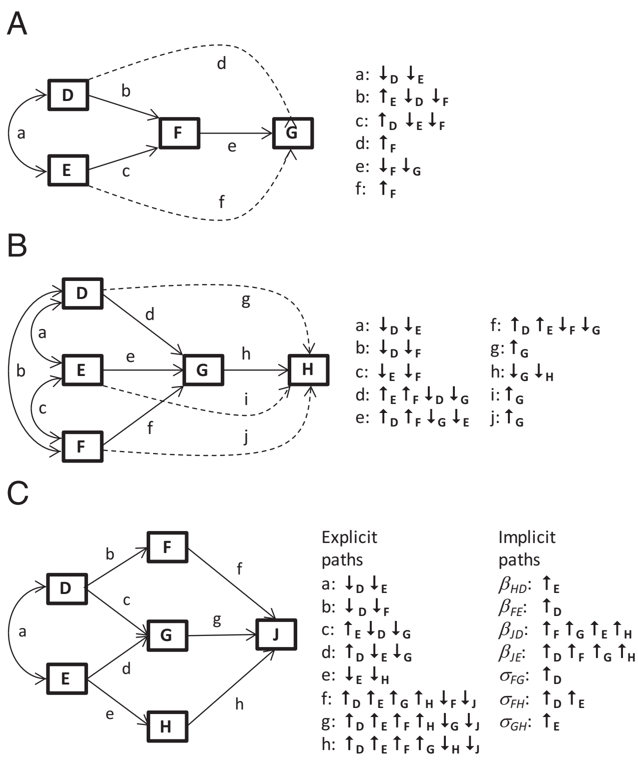

Figure 3. Increasingly complex models containing multiple fallible triangles and increasingly complicated patterns of over- and underestimation (represented by up and down arrows, respectively). Arrows pertain only to standardized path coefficients, as effects on unstandardized paths are more complex and vary with the magnitudes of certain covariances. A. Model 3a. B. Model 3b. C. Model 3c (the seven implicit paths are not depicted to avoid visual clutter).

Parameter Estimation

In this section, we unpack several principles that are generally true about the effect of model complexity on the inflation and attenuation of path coefficients. The first pertains to models that contain correlational paths between two or more exogenous manifest variables. For such correlational path coefficients, only underestimation is possible. When two manifest exogenous variables are simply correlated with each other, no other variable affects the path that connects them; consequently, their correlation cannot be affected by measurement error in any other variable in the model. Nevertheless, these correlational paths will be underestimated as a result of measurement error in either or both of these variables. One way that models can become increasingly complex is by inclusion of a larger number of purely exogenous variables, thereby increasing the number of correlational paths that are subject to underestimation. If we think of these correlations as two-variable combinations of k exogenous variables, then the number of potentially underestimated correlations equals the number of ways k variables can be grouped two-at-a-time: $C(k,2) = k! / (2!(k - 2)!) = k(k - 1) / 2$.

Second, directional paths potentially are subject to both under- and overestimation, depending on the model. Directional paths connect upstream variables to downstream variables. As models become more complex, the number of such directional paths is likely to increase. If m is the number of downstream variables and k is the number of purely upstream variables, then the total number of such directional paths in recursive models will equal km (the number of possible paths from the k upstream variables to the m downstream variables) plus $C(m,2)$, the number of two-variable connections among m downstream variables, for a total of $km + m(m - 1) / 2$ directional paths. In the interest of parsimony and theoretical relevance, actual models may not include all such paths; however, the *potential* for a larger number of directional paths clearly grows as either k or m increases.

Third, *underestimation* of any directional path coefficient occurs only because of unreliability in the two variables that are connected by the path in question (e.g., Bedeian, Day, & Kelloway, 1997). In general, let T_i represent one of q tracings between a particular pair of variables in a given model. For example, in Model 3c, five tracings connect variable G to variable J: $T_1 = g$, $T_2 = cbf$, $T_3 = deh$, $T_4 = caeh$, and $T_5 = dabf$. In general, the correlation between two variables like G and J will equal the sum of these tracings, $\rho_{GJ} = \sum_{i=1}^{q} T_i$. Here, we partition out the direct effect g from the rest of the $q - 1$ tracings:

[4] We make several assumptions in our structural model path diagrams: (a) variables that are vertically aligned are contemporaneous; (b) left-to-right movement assumes passage of time; (c) relations among contemporaneous variables are correlational; and (d) directional relations require the passage of time. These strictures dictate the kind of implicit or explicit path that can exist between any two variables.

$$\rho_{GJ} = g + \sum_{i=1}^{q-1} T_i. \tag{4}$$

If G and J are represented by fallible measures (G′ and J′), then their correlation underestimates ρ_{GJ} according to the classic attenuation formula, $\rho_{G'J'} = \rho_{GJ}\sqrt{\rho_{G'G'}\rho_{J'J'}}$, where $\rho_{G'G'}$ and $\rho_{J'J'}$ are the reliabilities of G′ and J′, respectively. Substituting Equation 4 into the attenuation formula yields

$$\rho_{G'J'} = (g + \sum_{i=1}^{q-1} T_i)\sqrt{\rho_{G'G'}\rho_{J'J'}} \tag{5}$$

or

$$\rho_{G'J'} = g\sqrt{\rho_{G'G'}\rho_{J'J'}} + \sum_{i=1}^{q-1} T_i\sqrt{\rho_{G'G'}\rho_{J'J'}}. \tag{6}$$

Clearly, the attenuating effects of unreliability in G′ and J′ are conveyed directly to path g (as well as the sum of the rest of the tracings between G′ and J′).

Fourth, the *overestimation* of any directional path is due to the underestimation of the sum of all valid tracings responsible for the covariation between the two variables of interest (other than the tracing that consists of the target path). For example, in Model 3c, *overestimation* of path g derives from underestimation of the other four tracings (*cbf*, *deh*, *caeh*, and *dabf*), the sum of which is $\sum_{i=1}^{q-1} T_i$. Let g′ and $\sum_{i=1}^{q-1} T_i'$ represent the expected values of g and $\sum_{i=1}^{q-1} T_i$ when any of the intermediate variables (i.e., D, E, F, and H but not G or J) in these tracings is measured with error. Therefore,

$$\rho_{GJ} = g' + \sum_{i=1}^{q-1} T_i'. \tag{7}$$

Substituting Equation 4 into Equation 7 gives us

$$g + \sum_{i=1}^{q-1} T_i = g' + \sum_{i=1}^{q-1} T_i' \tag{8}$$

and

$$g' - g = \sum_{i=1}^{q-1} T_i - \sum_{i=1}^{q-1} T_i'.$$

In other words, path g' will overestimate path g to the exact but opposite degree that unreliability in D, E, F, and H attenuates the sum of the other tracings that connect G to J (i.e., the degree to which $\sum_{i=1}^{q-1} T_i' \neq \sum_{i=1}^{q-1} T_i$).

Effect of Measurement Error on Goodness-of-Fit

One might ask whether estimating path coefficients while testing a relatively parsimonious model via structural equation modeling (SEM) would provide clues as to the existence of these problems and hints as to their resolution. In one small way, the answer is "yes," but in two large ways, the answer is "no." To explain, we first provide a verbal description of the diverse effects of measurement error on goodness-of-fit. Then we follow this with an example (coupled with Monte Carlo simulations) that supports these claims about the effect of measurement error.

The answer is a small "yes" in that increased measurement error exerts pressure to incorporate previously "missing" paths into the model, as demonstrated in our discussion of the inflationary effect above. When the coefficients for these paths are fixed at zero, this pressure is manifested in poor fit. The first big "no," however, derives from the somewhat disturbing fact that this pressure often derives from measurement unreliability, not necessarily from the existence of or need for paths connecting the underlying variables. For example, in Model 3c, pressure to include the D→J path derives from unreliability in E, F, G, and H, not necessarily from the existence of an effect of true D on true J. Inclusion of such paths may be necessary to achieve a good fit; however, if they are significant, these paths will likely be misinterpreted as support for a true D → true J effect, not as an accommodation of measurement unreliability in other intermediary variables. Even if researchers were aware of this possibility, they cannot easily know which paths are driven by measurement error and which are driven by substantive connections between the underlying variables. So, the better answer is "no," in that the degradation of model fit likely will be misinterpreted.

A second big "no" derives from the fact that pervasive measurement error can actually improve goodness-of-fit, making matters even more complicated. As measurement error increases, the total observed variance for each of the affected measures also rises; however, the covariances are unaffected. This attenuates the observed correlations among these variables (Duncan, 1975). As the observed correlations approach zero and discrepancies between the model and the data diminish, the implicit paths will be under less pressure to change; hence, conventional fit indices will provide evidence of improved fit. In other words, the model appears to provide a better fit to the data, not because the model explains more information but because the covariance matrix contains less information that has to be explained. When measurement error is pervasive, even serious model misspecification can go undetected. This problem pertains not only to manifest variable path analysis but to latent variable models as well, even when measurement error is accommodated by including unique factors in the measurement model (Hancock & Mueller, 2011; Heene et al., 2011; Lomax, 1986; Shevlin & Miles, 1998). Thus measurement error simultaneously results in both the increase and decrease in goodness-of-fit. The net result of these opposing forces is a function of model-specific characteristics and the degree of measurement error.

An Example

These general points become clearer with an example. Further, we can use this example to highlight how differential reliabilities in various parts of a model can lead to very different conclusions. Imagine a three-wave, longitudinal study designed to test two competing theories. Theory I stipulates that X affects Y and that this relation is partially mediated by A. Theory II also stipulates that X affects Y, but posits that B partially mediates the relation. Let us also consider Theory III, which stipulates that Theories I and II are *both* true. Based on this, a longitudinal study of mediation is conceived and analyses are planned as depicted in Model 4a (see Figure 4), in which subscripts for the variables denote the waves or time points in a longitudinal design. At a conceptual level, this is a good model. Following Cole and Maxwell's (2003) recommendations for longitudinal tests of mediation, each prediction of a dependent variable (A2, B2, and Y3) included the statistical control for prior levels of the same variable (A1, B1, and

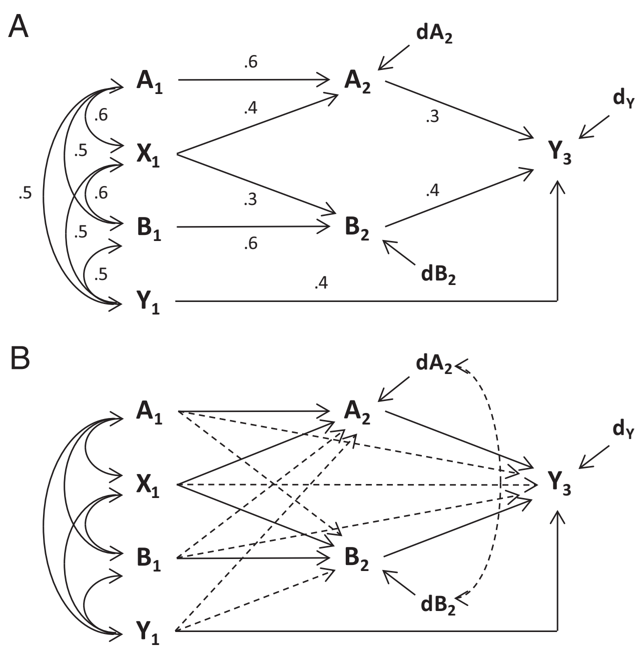

Figure 4. A. Model 4a reflects population path coefficients among true variables, with subscripts indicating waves or time points in a longitudinal design. B. Model 4b is a saturated version of Model 4a after adding all other plausible paths (i.e., the broken arrows), even though the population coefficients for these paths are zero when all variables are measured without error.

Y1, respectively). Regarding Theory I, the model allows estimation and testing of (a) the direct effect of X_1 on A_2 controlling for prior A_1, (b) the direct effect of A_2 on Y_3 controlling for prior Y_1 and B_2, and (c) the indirect effect of X_1 on Y_3 via the Time 2 mediator, A_2. Likewise for Theory II, the model can be used to estimate and test (a) the direct effect of X_1 on B_2 controlling for prior B_1, (b) the direct effect of B_2 on Y_3 controlling for prior Y_1 and A_2, and (c) the indirect effect of X_1 on Y_3 via the Time 2 mediator, B_2. Let us further imagine that Theory I and Theory II are equally valid and that population, standardized path coefficients would be those represented in the Figure 4 depiction of Model 4a, if all variables were measured perfectly. The indirect effect of $X_1 \rightarrow A_2 \rightarrow Y_3 = .4 * .3 = .12$ is equal to the indirect effect of $X_1 \rightarrow B_2 \rightarrow Y_3 = .3 * .4 = .12$. Furthermore, taken together the two mediators completely explain that part of the $X_1 \rightarrow Y_3$ relation not explained by Y_1 such that the $X_1 \rightarrow Y_3$ direct effect goes to zero when true A_2 and true B_2 are statistically controlled.

The Effects of Pervasive Unreliability on Model Fit

In this section we consider two broad circumstances under which pervasive unreliability affects model fit. The first is when the model is correctly specified. The second is when the model contains modest specification error. We use Model 4a as the basis for both circumstances.

A correctly specified model. First, we assume that the fixed and free parameters in Model 4a represent true relations among the underlying variables. That is, all fixed paths are truly zero in the population. Admittedly this represents an idealized (and unrealistic) situation, as all models are misspecified to some degree (MacCallum, 2003); however, witnessing the effect of measurement error on such an idealized model is informative.

As described above, we anticipated that the effects of measurement error on model fit would be complex. On one hand, fixing path coefficients to zero even when they truly are zero will create evidence of model misfit as measurement error pervades a set of variables, as this measurement error can put pressure on fixed paths to deviate from zero. On the other hand, pervasive measurement error will also diminish the observed correlations, therefore limiting the size of the discrepancies between the observed correlations and those implied by the model.[5] As these discrepancies lie at the core of almost all goodness-of-fit indices, their reduction will enable the model (really any model) to fit the data well. To observe the net results of these opposing effects, we conducted a Monte Carlo simulation examining the effects of pervasive measurement error on goodness-of-fit in Model 4a at 11 different levels of reliability, ranging from 1.00 down to 0.00.

For the first level (where reliabilities for all measures were 1.00), we computed a population covariance matrix based on the path coefficients in Model 4a (see the first seven variables in Table 2). Using Mplus 6.12, we generated 1,000 multivariate normal samples, each with 200 observations, to which we then fit Model 4a. For each test, we recorded the *p* value associated with the chi-square statistic. Across these 1,000 repetitions, we calculated the probability that the *p* value was less than .05 (i.e., the probability that the model generated a statistically poor fit). The result was a value of .054, very close to the nominal alpha. This value is plotted in Figure 5 above the *x*-axis category for reliability = 1.0. The next 10 simulations were identical to the first except that the population variances and correlations were changed to reflect diminishing levels of reliability (i.e., .90, .80, .70, ...00). Within each simulation, we assumed equal reliabilities: $\rho_{A_1'A_1'} = \rho_{B_1'B_1'} = \rho_{X_1'X_1'} = \rho_{Y_1'Y_1'} = \rho_{A_2'A_2'} = \rho_{B_2'B_2'} = \rho_{Y_3'Y_3'} = \rho_{..}$. The results are plotted in Figure 5.

The findings are striking. Evidence of two competing trends emerged. First, as reliability dropped from 1.0 to .50, the likelihood of rejecting Model 4a *rose* from .054 to .931, despite the fact that Model 4a is the true model when there is no measurement error. Even with reliabilities of .90, the likelihood of rejecting the model was quite high (.391). Second, as reliability continued to drop from .50 to zero, a second trend predominated; the likelihood of rejecting Model 4a *dropped* from .931 down to .074. We interpret this curve as evidence of two competing processes. At relatively high levels of reliability, anything less than perfect reliability requires the inflation of various path coefficients; however, when these path coefficients are fixed to zero, this inflationary pressure translates into model misfit. At lower levels of reliability, a second process prevails: the overall diminution of the observed correlations makes the detection of misfit less and less possible. Our overarching conclusion is that goodness-of-fit information will be misleading (in one way or another), as even modest

[5] This pertains to models in which overidentification is achieved by fixing some path coefficients to zero or by imposing various equality constraints. This may not pertain in the relatively rare case where a model is overidentified by fixing some path coefficients to values that are not zero.

Table 2
Population Correlations Among Variables With and Without Measurement Error for Models 4a and 4b

	Variables without measurement error							Variables with measurement error						
Variable	A_1	X_1	B_1	Y_1	A_2	B_2	Y_3	A_1'	X_1'	B_1'	Y_1'	A_2'	B_2'	Y_3'
A_1	1.000													
X_1	.600	1.000												
B_1	.500	.600	1.000											
Y_1	.500	.500	.500	1.000										
A_2	.840	.760	.540	.500	1.000									
B_2	.480	.660	.780	.450	.552	1.000								
Y_3	.644	.692	.674	.730	.721	.746	1.000							
A_1'	.700	.420	.350	.350	.588	.336	.451	1.000						
X_1'	.420	.700	.420	.350	.532	.462	.484	.294	1.000					
B_1'	.350	.420	.700	.350	.378	.546	.472	.245	.294	1.000				
Y_1'	.350	.350	.350	.700	.350	.315	.511	.245	.245	.245	1.000			
A_2'	.588	.532	.378	.350	.700	.386	.505	.412	.372	.265	.245	1.000		
B_2'	.336	.462	.546	.315	.386	.700	.522	.235	.323	.382	.221	.270	1.000	
Y_3'	.451	.484	.472	.511	.505	.522	.700	.316	.339	.330	.358	.353	.365	1.000

Note. Population variances for A_1–Y_3 are 1.0; population variances for A_1'–Y_3' are ~2.04.

amounts of measurement error permeate a properly specified manifest variable path analysis.

A misspecified model. Second, we assumed that the model does *not* precisely reflect the actual relations among the underlying variables; that is, the model is misspecified. Specifically, at least one fixed path is not truly zero in the population. Under this more realistic scenario (Cudeck & Browne, 1992; Tucker, Koopman, & Linn, 1969), we would expect to discover statistical evidence of poor fit, assuming that we have a large enough sample size. The effects of measurement error on the power to detect such specification error are illuminating.

To demonstrate, we retained the same data-generating model; however, we modified the fitted Model 4a so that the $Y_1 \rightarrow Y_3$ path was constrained to zero. We again constructed a set of 11 covariance matrices based on the original Model 4a, each reflecting a different amount of measurement error. As before, we assumed $\rho_{A_1'A_1'} = \rho_{B_1'B_1'} = \rho_{X_1'X_1'} = \rho_{Y_1'Y_1'} = \rho_{A_2'A_2'} = \rho_{B_2'B_2'} = \rho_{Y_3'Y_3'} = \rho_{..}$. We applied the same Monte Carlo methods described above and obtained the results depicted in Figure 5b. For higher levels of reliability (>0.40), the power to detect the specification error was quite good. As reliability diminished, however, the power to detect the model misspecification dropped precipitously.

The Effect of Pervasive Measurement Error on Power to Detect Parameter Coefficients

Based on the results in Figure 5a, we would anticipate that researchers who use measures that contain modest amounts of error are likely to be faced with evidence of model misfit. Rightly or wrongly, let us assume that such evidence would lead researchers to add theoretically justifiable paths in order to achieve an acceptable model. Instead of considering all combinations of theoretically justifiable model modifications, we elected to add directional paths from *all* upstream variables to *all* downstream variables, as depicted in Model 4b in Figure 4. (To be clear, we do not advocate this data analytic strategy but implement it here to witness key effects in the context of a well-fitting model.) These additional paths (plus the correlation between the disturbances for A2 and B2) saturate the model, making goodness-of-fit irrelevant and providing us the opportunity to examine the coefficients not only for all of the original paths but also for the plausible post hoc paths as well. Following the methods described above, we conducted 11 Monte Carlo simulations, one for each of 11 levels of reliability (assuming $\rho_{A_1'A_1'} = \rho_{B_1'B_1'} = \rho_{X_1'X_1'} = \rho_{Y_1'Y_1'} = \rho_{A_2'A_2'} = \rho_{B_2'B_2'} = \rho_{Y_3'Y_3'} = \rho_{..}$). In each, we drew 1,000 multivariate normal samples, each with $N = 200$.

Let us consider Theory III to be represented by the seven solid arrows in Model 4b. Within the constraints of this model, one could argue that support accrues to Theory III as more and more of the key seven paths prove to be statistically significant. To witness the effect of measurement error on this approach, we set alpha to .05 for each path coefficient and estimated the joint probability that all seven paths would be statistically significant under conditions of diminishing measurement reliability. These results are represented by Line 7 in the power plot of Figure 6. When $\rho_{..} = 1.0$, power to detect all seven paths was very high (.974); however, power diminished appreciably with reliability. For reliabilities of .9, .8, and .7, power was .859, .715, and .589, respectively (as we slide down Line 7 from right to left), diminishing as reliability was reduced. We also calculated power to detect partial support for the theory. Specifically, we calculated the joint probabilities to detect at least six out of seven paths, at least five out of seven paths, etc., down to at least one path. Those results are plotted in Lines 1–6 to the left of Line 7 in Figure 6. Of course, power for these scenarios was greater, but all evinced the same basic pattern shown for Line 7.

Let us now regard the seven broken directional arrows as representing relations that are contrary to Theory III. In the context of Model 4b, one might argue that support for Theory III diminishes as more and more of these seven broken arrows proves to be significant. In this example, these seven path coefficients truly would be zero if all variables were measured without error. For example, if the truly zero $X_1 \rightarrow Y_3$ path were significant, the theoretical tenet that A_2 and B_2 completely mediated the $X_1 \rightarrow Y_3$ relation would be spuriously rejected. To examine the effect of

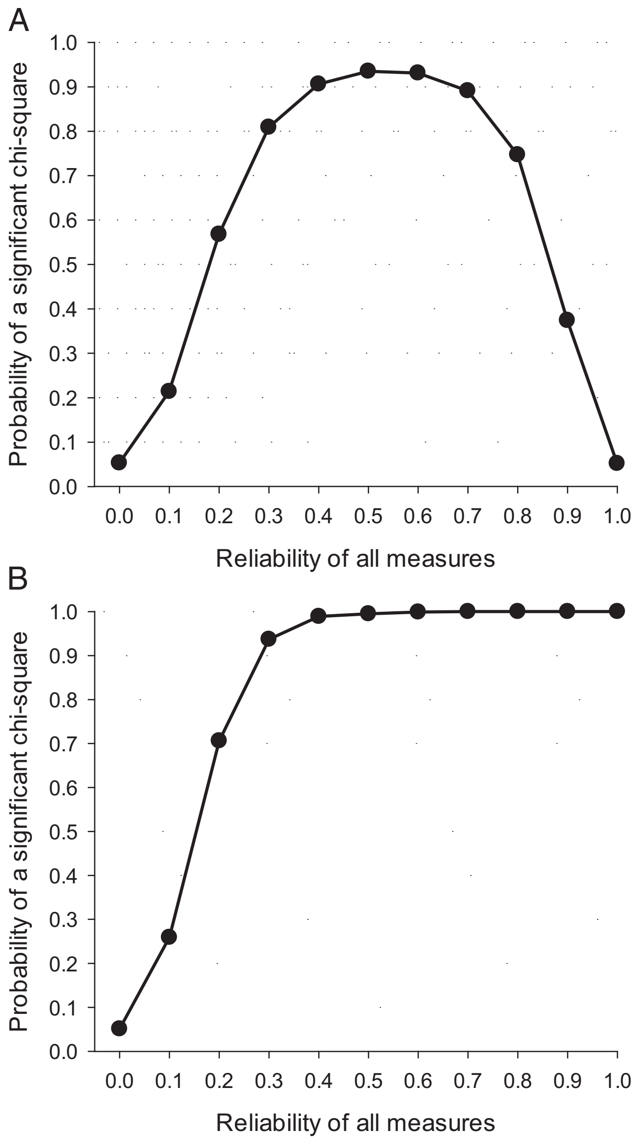

Figure 5. A. Upper plot is for the probability of model misfit (i.e., a significant chi-square statistic) for Model 4a tested against 7 × 7 covariance matrices attenuated because of varying degrees of (un)reliability. B. The second plot shows the probability of a significant chi-square statistic for a modified Model 4a in which the $Y_1 \rightarrow Y_3$ path was constrained to zero. Thus, A represents Type I error, and B represents power.

measurement error on this aspect of the theory, we assessed the probability that any of the seven truly nonzero paths would be erroneously detected as significant under conditions of diminishing measurement reliability. For each level of reliability, we plotted the joint probability that at least one of the seven truly zero paths was significant (see Line 1 in the Type I error plot in Figure 6). Interestingly, this line mirrors the goodness-of-fit curve in Figure 5. As reliability drops from 1.0 to around .5, the familywise Type I error rate increases, reflecting the effect of measurement error on the overestimation of some truly zero path coefficients. As reliability drops from .5 to 0, the familywise Type I error rate decreases, reflecting the generic attenuation effect of measurement error on the observed correlations. We also calculated the familywise Type I error rates for at least two paths, at least three paths, etc., up to all seven paths. Those lines appear beneath the first one in the Figure 6 Type I error plot.

The conclusion to be drawn from the two graphs in Figure 6 is twofold. First, even modest amounts of measurement error (e.g., reliabilities of 0.8) can generate nontrivial levels of Type II error, resulting in a failure to support all aspects of a completely valid theory. Second, it takes even less measurement error (e.g., reliabilities of 0.9) to generate high levels of Type I error, leading to the spurious "discovery" of relations that do not truly exist when the variables are measured without error. Following the guidelines recommended by Nunnally and Bernstein (1994), many researchers conduct path analysis (and multiple regression) when reliabilities are only as large as 0.80. We contend that such "high" reliabilities may actually be too low.

The Effects of Differential Reliability on Parameter Estimation

In the previous demonstrations, we assumed that measurement error was evenly distributed throughout the model. Continuing with our Model 4a example, we now demonstrate how *differential* reliability in various parts of a model can lead to very different conclusions in ways that are not easily anticipated (James et al.,

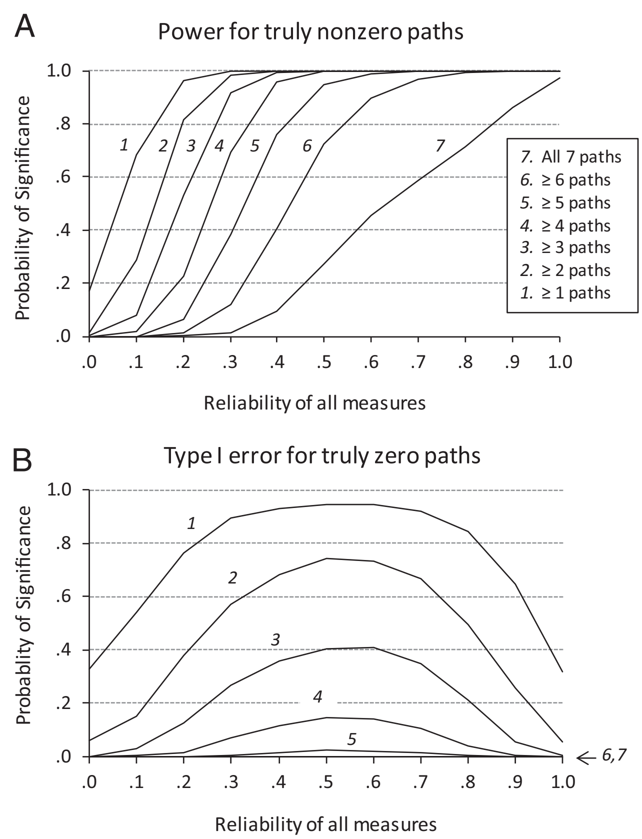

Figure 6. A. The probabilities of showing significance for truly nonzero paths (power). B. The probabilities of showing significance for truly zero paths (Type I error) in Model 4b.

1982; Kenny, 1979). We begin by assuming again that Model 4a is the true model and that the path coefficients in Figure 4 represent the population parameters when all variables are measured without error. The population correlations among the error-free variables, A_1, X_1, B_1, Y_1, A_2, B_2, and Y_3, are represented in Table 2. If one tested the model using these measures, the model would fit perfectly and the path coefficients in Model 4a would be completely recovered. Alternatively, if one had an imperfect measure of every construct in the model (such that each measure correlated 0.7 with its construct), then the expected correlations would be those displayed in Table 2 for the variables A_1', X_1', B_1', Y_1', A_2', B_2', and Y_3'.[6]

First, we examined the effects of measurement error in only one variable at a time. That is, we calculated the expected values for all parameter estimates when six variables were measured perfectly and only one variable was measured with error. As in the previous section, we based these calculations on analyses of the fully saturated Model 4b to prevent bias due to model misspecification. These results appear in Table 3. Boldface values are over- or underestimates relative to what the standardized coefficients would have been had the variables been measured perfectly. Several key findings emerged.

1. Measurement error in any single variable caused inflation or attenuation of multiple parameters. Depending on which variable contained measurement error, the number of affected parameters ranged from three to 12 (including the correlation between the disturbance terms for A_2 and B_2).

2. Measurement error in any particular variable always resulted in the attenuation of paths that involved that variable, unless the true coefficient for that path was already zero.

3. In almost every case, fallibility in even one variable resulted in the inflation of a path coefficient that was truly zero, a condition that would have suggested the spurious addition of at least one path to the original model. (The exception was when Y_3 was the fallible variable because a purely endogenous variable cannot explain correlations among exogenous variables.)

4. The absolute sizes of the over- and underestimations were often quite large, depending on the fallible variable and the target path. In Table 3, relatively large deviations (e.g., ≥ 0.15) occurred 25 times.

5. For some of the directional paths, the direction of bias switched depending on which variable was fallible (e.g., the $X_1 \rightarrow A_2$ and the $X_1 \rightarrow B_2$ paths).

6. Depending on which variable was fallible, erroneous estimation of the indirect effects led to very different theoretical conclusions: (a) Support for Theory I could be either spuriously enhanced or spuriously diminished by over- or underestimating the indirect effect through A_2, while leaving the indirect effect through B_2 untouched. (b) Support for Theory II could be either spuriously enhanced or diminished by over- or underestimating the indirect effect through B_2, while leaving the indirect effect through A_2 unaffected. (c) Support for both theories could be diminished by underestimation of the indirect effects through both A_2 and B_2. Only when Y_1 was the fallible variable was support for both indirect effects unbiased (because in this model Y_1 has no direct effect on either the A_2 or B_2 mediator).

7. Saturating the model with all theoretically defensible paths failed to eliminate problems of over- and underestimation.

Second, we examined the over- and underestimation that resulted when two variables contained measurement error. We focused on six two-variable combinations (A_2 and B_2, X_1 and Y_3, A_1 and Y_1, B_1 and Y_1, A_1 and B_2, B_1 and A_2) instead of testing all 21 combinations. The results were again dramatic (see Table 4).

1. Altogether more parameters were misestimated than when only one variable was fallible. Across the seven examples, the number of affected parameters ranged from three to 15 (including the correlation between the disturbance terms for A_2 and B_2).

2. Among the affected parameters, instances of underestimation were greater in Table 4 than in Table 3 (where only one variable was fallible).

3. Nevertheless, every case resulted in multiple instances in which a truly zero path coefficient was overestimated.

4. The magnitude of the distortion was often quite large. In Table 4, relatively large distortions (e.g., $\leq -.15$ or $\geq .15$) occurred 39 times.

5. Impact on the indirect effects was profound. In the first two cases, both indirect effects were substantially underestimated, virtually eliminating any support for either Theory I or Theory II. In the next two cases, one indirect effect was substantially overestimated while the other remained unaffected. In the last two cases, one indirect effect was overestimated while the other was underestimated, generating spurious support for one theory over the other, purely as a function of which variables contained error.

Finally, we considered cases where three or more variables contained measurement error. We focused on four possibilities: all endogenous variables (A_2, B_2, Y_3) were fallible, all exogenous variables (A_1, B_1, X_1, Y_1) were fallible, a mix of exogenous and endogenous variables (X_1, A_2, B_2, Y_3) were fallible, and all variables were fallible. The results appear in Table 5.

[6] The points made in this section do not require that the correspondence of manifest variables to their underlying constructs literally be as high as 1.0 or as low as .7; however, both scenarios occur in many published models. Variables such as grade, age, sex, and group membership have near perfect reliabilities and are often included in path analyses as perfect measures (e.g., D. W. King, King, & Foy, 1996; Taasoobshirazi & Carr, 2009). Conversely, manifest variables often have factor loadings that are .7 or less (e.g., Bentler & Speckart, 1981; South, Krueger, & Iacono, 2011).

Table 3
Bias in SEM Coefficients for Model 4b When Reliability Is Diminished One Variable at a Time

Parameter	Parameter value	Bias when these variables are measured with error						
		Only A_1	Only B_1	Only X_1	Only Y_1	Only A_2	Only B_2	Only Y_3
Original paths								
cov A_1,X_1	.60	**−.18**	.00	**−.18**	.00	.00	.00	.00
cov A_1,B_1	.50	**−.15**	**−.15**	.00	.00	.00	.00	.00
cov A_1,Y_1	.50	**−.15**	.00	.00	**−.15**	.00	.00	.00
cov X_1,B_1	.60	.00	**−.18**	**−.18**	.00	.00	.00	.00
cov X_1,Y_1	.50	.00	.00	**−.15**	**−.15**	.00	.00	.00
cov B_1,Y_1	.50	.00	**−.15**	.00	**−.15**	.00	.00	.00
$X_1 \rightarrow A_2$	.40	**.15**	.00	**−.21**	.00	**−.12**	.00	.00
$A_1 \rightarrow A_2$	.60	**−.30**	.00	**.09**	.00	**−.18**	.00	.00
$B_1 \rightarrow B_2$	.60	.00	**−.30**	**.07**	.00	.00	**−.18**	.00
$X_1 \rightarrow B_2$	.30	.00	**.15**	**−.16**	.00	.00	**−.09**	.00
$Y_1 \rightarrow Y_3$	.40	.00	.00	.00	**−.18**	.00	.00	**−.12**
$A_2 \rightarrow Y_3$	.30	.00	.00	.00	.00	**−.23**	.00	**−.09**
$B_2 \rightarrow Y_3$	.40	.00	.00	.00	.00	.00	**−.26**	**−.12**
Added paths								
$A_1 \rightarrow B_2$	.00	.00	**.06**	**.07**	.00	.00	.00	.00
$A_1 \rightarrow Y_3$	.00	.00	.00	.00	**.06**	**.15**	.00	.00
$B_1 \rightarrow A_2$	.00	**.06**	.00	**.09**	.00	.00	.00	.00
$B_1 \rightarrow Y_3$	.00	.00	.00	.00	**.06**	.00	**.18**	.00
$X_1 \rightarrow Y_3$	.00	.00	.00	.00	**.05**	**.10**	**.09**	.00
$Y_1 \rightarrow A_2$	.00	**.09**	.00	**.04**	.00	.00	.00	.00
$Y_1 \rightarrow B_2$	.00	.00	**.09**	**.03**	.00	.00	.00	.00
cov dA_2,dB_2	.00	.00	.00	**.04**	.00	.00	.00	.00
Indirect effects								
$X_1 \rightarrow A_2 \rightarrow Y_3$	.12	**.05**	.00	**−.06**	.00	**−.10**	.00	**−.04**
$X_1 \rightarrow B_2 \rightarrow Y_3$	.12	.00	**.06**	**−.06**	.00	.00	**−.09**	**−.04**

Note. SEM = structural equation modeling. Boldface signifies nonzero bias from the parameter values shown in Figure 4a.

1. Misleading results pervaded all examples. Between 10 and 21 parameters were distorted in any given example. In two examples, every path was affected. In Table 5, large distortions occurred 33 times.

2. The great majority of the original (nonzero) paths were underestimated.

3. Almost all of the truly zero paths were overestimated, a situation that could mislead investigators into adding theoretically spurious paths to the original model.

4. Including these paths did not even begin to eliminate the distortions of the original path coefficients.

Discussion, Recommendations, and Conclusions

This article demonstrates that five potentially serious problems can arise from the use of fallible measures in manifest variable path analysis. First, as measurement error pervades a given data set, virtually all path coefficients will be under- or overestimated. Second, even a little measurement error can cause valid models to appear invalid. Third, when large amounts of measurement error pervade a model, power diminishes, reducing the chance of rejecting an invalid model. Fourth, differential measurement error in various parts of a model can change the substantive conclusions that derive from path analysis. Fifth, all of these problems become increasingly serious and intractable as models become more complex. In this section, we briefly elaborate each of these problems and then discuss a series of approaches for coping with them.

The first issue is that measurement error can result in the over- or underestimation of path coefficients almost anywhere in a manifest variable path analysis. The direction of bias depends on the location of measurement error in the model. Unreliability among exogenous variables will attenuate their correlation, reducing power. Unreliability in either an upstream or downstream variable will attenuate the standardized coefficient for the directional path between them, also reducing power and increasing the possibility of Type II error. Unreliability in either a predictor or mediator that is part of a fallible triangle will inflate the path between the other two variables. Depending upon the model and sample size, truly negligible and nonsignificant path coefficients can become substantial and statistically significant. A useful avenue for future research will involve investigating the magnitudes of these effects in simple and complex models.

The second issue is that relatively small amounts of measurement error can cause valid models to appear invalid. This happens when a path that is truly zero is fixed to zero, but unreliability in a “third” variable elsewhere in the model requires the inflation of this path coefficient. The third variable is essentially a control variable. When measured perfectly, the third variable explains the relation between the other two variables, but when measured imperfectly, it does not and fixing the path between the other two variables to zero then generates evidence of misfit. The magnitude of this effect can be profound. In one example, we show that a drop

Table 4
Bias in SEM Coefficients for Model 4b When Reliability Is Diminished Two Variables at a Time

Parameter	Parameter value	Bias when these variables are measured with error					
		A_2, B_2	X_1, Y_3	A_1, Y_1	B_1, Y_1	A_1, B_2	B_1, A_2
Original paths							
cov A_1,X_1	.60	.00	**−.18**	**−.18**	.00	**−.18**	.00
cov A_1,B_1	.50	.00	.00	**−.15**	**−.15**	**−.15**	**−.15**
cov A_1,Y_1	.50	.00	.00	**−.25**	**−.15**	**−.15**	.00
cov X_1,B_1	.60	.00	**−.18**	.00	**−.18**	.00	**−.18**
cov X_1,Y_1	.50	.00	**−.15**	**−.15**	**−.15**	.00	.00
cov B_1,Y_1	.50	.00	.00	**−.15**	**−.25**	.00	**−.15**
$X_1 \rightarrow A_2$	.40	**−.12**	**−.21**	**.17**	.00	**.15**	.00
$A_1 \rightarrow A_2$	.60	**−.18**	**.09**	**−.29**	.00	**−.30**	.00
$B_1 \rightarrow B_2$	.60	**−.18**	**.07**	.00	**−.29**	**−.18**	.00
$X_1 \rightarrow B_2$	.30	**−.09**	**−.16**	.00	**.17**	**−.09**	.00
$Y_1 \rightarrow Y_3$	.40	.00	**−.12**	−.18	**−.18**	.00	.00
$A_2 \rightarrow Y_3$	.30	**−.23**	**−.09**	.04	.00	.00	.00
$B_2 \rightarrow Y_3$	.40	**−.26**	**−.12**	.00	**.03**	**−.26**	.00
Added paths							
$A_1 \rightarrow B_2$	.00	.00	**.07**	.00	**.07**	.00	.00
$A_1 \rightarrow Y_3$	.00	**.15**	.00	**.02**	**.07**	.00	.00
$B_1 \rightarrow A_2$	.00	.00	**.09**	**.07**	.00	**.06**	.00
$B_1 \rightarrow Y_3$	.00	**.18**	.00	**.07**	**.02**	**.18**	.00
$X_1 \rightarrow Y_3$	.00	**.19**	.00	**.04**	**.05**	**.09**	.00
$Y_1 \rightarrow A_2$	.00	.00	**.04**	**.05**	.00	**.09**	.00
$Y_1 \rightarrow B_2$	.00	.00	**.03**	.00	**.05**	.00	.00
cov dA_2,dB_2	.00	.00	**.04**	.00	.00	.00	.00
Indirect effects							
$X_1 \rightarrow A_2 \rightarrow Y_3$	.12	**−.10**	**−.08**	**.08**	.00	**.05**	.00
$X_1 \rightarrow B_2 \rightarrow Y_3$	.12	**−.09**	**−.08**	.00	**.08**	**−.09**	.00

Note. SEM = structural equation modeling. Boldface signifies nonzero bias from the parameter values shown in Figure 4a.

in reliability from 1.0 to .8 increases the chances of model rejection from 5% to 75%. Future research is needed to examine the severity of these effects in a wider variety of models.

Conversely, the third issue is that pervasive measurement error can also reduce power to reject a truly misspecified model. Measurement error generally has the effect of attenuating the observed correlations. When observed correlations are small, the discrepancies between them and the model-implied correlations also will be small, making it difficult to reject almost any model. In the current example, this effect was strong at only very low levels of reliability. Another direction for future research will be exploration of the extent to which small, medium, and large specification errors may go undetected at various levels of measurement error.

The fourth issue is that different degrees of measurement error in different parts of a model can change the substantive conclusions that derive from manifest variable path analysis. We found that measurement unreliability in even one or two variables can bias relatively few path coefficients or a majority of path coefficients, depending on where in the model the fallible variables reside. The direction of this bias can vary considerably, either overestimating or underestimating specific path coefficients, depending on which variables are fallible. Furthermore, different patterns of reliability can even generate spurious evidence of one theory over another.

The fifth issue is that these problems become more likely, more serious, and less predictable as models become more complex. As more variables are added to a model, more paths must be either constrained or estimated, actions that increase the opportunity for measurement error to bias path coefficients and affect model fit. As models become more complex, the number of affected paths increases, the magnitude of the bias grows, and the net result of these effects becomes harder to anticipate.

Recommendations

Given that the problems described above can be quite serious (depending on the model and the degree of measurement error), we briefly present three broad recommendations of methods to reduce the impact of measurement error on manifest variable path analyses: using error reduction strategies, utilizing model-based error correction methods, and focusing on simpler models. Some of these approaches take multiple forms, and all carry with them their own shortcomings.

Error reduction strategies. Methods for reducing measurement error are well documented. One general strategy is to use longer assessments. Assuming parallel units, measures with more items tend to have higher reliabilities (Brown, 1910; Spearman, 1910). This logic pertains to the addition of items to tests, sampling behavior over a longer period of time, pooling across more observers, observing participants across multiple situations, and extending longitudinal studies across more time points. In practice, however, several practical limitations may curtail the theoretical advantages of these methods: (a) the best questionnaire items may already be part of the original measure, such that new items may

Table 5

Bias in SEM Coefficients for Model 4b When Reliability Is Diminished 3–7 Variables at a Time

Parameter	Parameter value	Bias when these variables are measured with error			
		A_2, B_2, Y_3	A_1, X_1, B_1, Y_1	X_1, A_2, B_2, Y_3	All seven
Original paths					
cov A_1,X_1	.60	.00	**−.31**	**−.18**	**−.31**
cov A_1,B_1	.50	.00	**−.25**	.00	**−.25**
cov A_1,Y_1	.50	.00	**−.25**	.00	**−.25**
cov X_1,B_1	.60	.00	**−.31**	**−.18**	**−.31**
cov X_1,Y_1	.50	.00	**−.25**	**−.15**	**−.25**
cov B_1,Y_1	.50	.00	**−.25**	.00	**−.25**
$X_1 \rightarrow A_2$	.40	**−.12**	**−.07**	**−.27**	**−.17**
$A_1 \rightarrow A_2$	.60	**−.18**	**−.18**	**−.11**	**−.30**
$B_1 \rightarrow B_2$	.60	**−.18**	**−.20**	**−.13**	**−.32**
$X_1 \rightarrow B_2$	.30	**−.09**	**−.02**	**−.20**	**−.11**
$Y_1 \rightarrow Y_3$	.40	**−.12**	**−.17**	**−.11**	**−.20**
$A_2 \rightarrow Y_3$	.30	**−.25**	**.07**	**−.23**	**−.16**
$B_2 \rightarrow Y_3$	.40	**−.30**	**.04**	**−.29**	**−.22**
Added paths					
$A_1 \rightarrow B_2$	.00	.00	**.13**	**.05**	**.09**
$A_1 \rightarrow Y_3$	.00	**.11**	**.02**	**.13**	**.10**
$B_1 \rightarrow A_2$	.00	.00	**.14**	**.07**	**.10**
$B_1 \rightarrow Y_3$	.00	**.13**	**.03**	**.15**	**.12**
$X_1 \rightarrow Y_3$	.00	**.13**	**.02**	**.06**	**.12**
$Y_1 \rightarrow A_2$	.00	.00	**.13**	**.03**	**.09**
$Y_1 \rightarrow B_2$	.00	.00	**.12**	**.02**	**.08**
cov dA_2,dB_2	.00	.00	**.14**	**.02**	**.07**
Indirect effects					
$X_1 \rightarrow A_2 \rightarrow Y_3$	.12	**−.11**	.00	**−.11**	**−.09**
$X_1 \rightarrow B_2 \rightarrow Y_3$	.12	**−.10**	.00	**−.11**	**−.09**

Note. SEM = structural equation modeling. Boldface signifies nonzero bias from the parameter values shown in Figure 4a.

not be as reliable as the original ones; (b) lengthening a measure may introduce new, unwanted variance due to fatigue, resentment, or loss of vigilance; (c) adding more time points can extend a study to a point in time when the targeted trend begins to change qualitatively; (d) changing to a more reliable method of measurement unintentionally changes the nature of the underlying construct. Clearly, lengthening or changing measures should be accompanied by sound psychometric support.

Rather than settling for a large number of potentially unreliable items and risk encountering the limitations noted above, an alternative to using longer assessment instruments is to generate better items. Assessment instruments are often of poor quality, and frequently do not undergo the multiple rounds of rigorous psychometric evaluation and scale development advocated by psychometricians. When it comes to maximizing parsimony and validity, using a smaller pool of highly reliable items is clearly preferable to using a larger pool of questionably reliable items.

Alternatively, when highly reliable measures are simply not available, the researcher may be able to obtain multiple measures (of the same or lower reliability) of each construct, which can be utilized in some type of latent variable analysis (e.g., structural equation modeling). The benefits of adopting a latent variable strategy are numerous and well known (Bollen, 1989; Kenny, 1979); however, in such designs, measurement continues to matter. Research has shown that the researcher is still well advised to choose the most reliable, valid, and domain-representative measures possible (Bedeian et al., 1997; Hancock & Mueller, 2011; Ledgerwood & Shrout, 2011; Tomarken & Waller, 2003). Furthermore, having more indicators per latent variable is better than having fewer (Kano, 2007; Marsh, Hau, Balla, & Grayson, 1998; Mulaik, 2009), and retaining relatively unreliable indicators is often preferable to discarding them (Bollen, 1989, p. 165; McCallum, 1972; Wickens, 1972). Indeed, Mulaik (2009) advocated including as many indicators as possible for each construct (with at least four per latent variable) to maximize construct validity and isolate the desired construct.

Despite the advantages associated with obtaining more measures, complications and dangers often emerge in practice. First, having multiple measures of multiple constructs increases the likelihood that particular pairs of measures will share method variance. Depending upon the measurement design, statistical controls for such problems may be impossible or (worse yet) highly misleading (Cole, Ciesla, & Steiger, 2007; Saris & Aalberts, 2003). Second, adding suboptimal measures to a study can lead inadvertently to construct validity problems. New measures may cross-load onto unintended latent variables or change qualitatively the nature of the intended latent variable (Cole et al., 2007). All things being equal, more is indeed better, but all things are not always equal.

Error correction strategies. In general, error reduction is better than error correction; however, sometimes constraints on time and financial resources can impede implementation of error reduction methods. In such cases, various post hoc methods to adjust for measurement error can be implemented. Here we discuss four: data-based correction for unreliability, model-based correction for unreliability, use of items as indicators, and parceling.

First, data-based correction for unreliability is possible when the researcher has only one indicator per construct but has estimates of the reliabilities of these indicators. These reliability estimates can be used to adjust the variance-covariance matrix for unreliability. Actually, this correction only affects the variances, not the covariances (Blalock, 1964; Duncan, 1975; Kenny, 1979).[7] The path analytic model is then fit to the original covariances and adjusted variances.

Second, model-based corrections are possible by using a latent variable structural equation model to conduct the manifest variable path analysis. In such models, one manifest variable represents each "latent" variable. For each manifest variable, the factor loading is set to 1.0, and the unique variance is set to a value based on knowledge about data quality (see Adcock, 1878; Bollen, 1989; Deming, 1943; Hayduk, 1987; Koopmans, 1937; Kummell, 1879; Williams & O'Boyle, 2008). Specifically, Hayduk (1987) recommended fixing the unique variance to a value dictated by an estimate of the reliability: $(1 - \hat{\rho}_{XX'})\sigma^2_X$.[8] In this way, the unique factor is interpretable as the unreliable part of the manifest variable, and the latent variable becomes the reliable part, theoretically rendering the latent variable error-free. Taking this approach further, DeShon (1998) advocated first computing a generalizability coefficient, an estimate of reliability based on multiple sources of measurement error (e.g., person, rater, item, and occasion). Then the unique variance and loading of a single, standardized indicator are fixed to functions of this coefficient, and the factor variance is fixed to 1.

Interestingly, in many cases, the data-based and model-based methods are equivalent (Rock, Werts, Linn, & Jöreskog, 1977; Werts & Linn, 1970). Potential advantages associated with these approaches are: (a) they allow the inclusion of prior knowledge about measurement quality; (b) they do not assume that the indicators are perfectly reliable; and (c) they reduce the typical effect of measurement error on the inflation and attenuation of structural coefficients. Unfortunately, these methods also bring with them several shortcomings: (a) Reliability estimates for the measures may not be available for the sample under investigation, and using estimates from one population to correct for error in another can yield spurious results. (b) Even when obtained from the same sample, reliability estimates are still only estimates. In any given study, they will over- or underestimate the actual reliability, giving rise to under- or overcorrections for measurement error. Further, treating any parameter as known, when it is not, can artificially bias standard errors of path coefficients toward zero, leading to increased Type I error rates. (c) Multiple sources of measurement error exist (e.g., error due to rater, item, occasion). Correcting for only one source of measurement error limits the generalizability of results (DeShon, 1998). (d) In the common factor model, the unique variance associated with an indicator is actually a combination of reliable, indicator-specific variance and error variance. Clearly, some contributors to unique variance (e.g., method variance) should not be considered part of the reliable variance of an indicator. But if at least some of the unique variance should be considered reliable, then treating the unique variance as if it consists entirely of error variance implies that the variable is less reliable than it really is, which can result in biased structural coefficients (Coffman & MacCallum, 2005). Although not a shortcoming of error correction strategies per se, it should be borne in mind that correcting for unreliability does not guarantee validity. Such corrections yield reliable indicators of whatever it is that the indicators measure, which is rarely isomorphic with the construct of interest.

A third model-based method is to use items as indicators. Many times a single measure is an aggregation of smaller units. Under some circumstances, these smaller units can serve as multiple indicators of a latent variable. Using disaggregated items as indicators has some advantages: (a) analyses are more likely to converge on proper solutions (Marsh et al., 1998); (b) additional factors are less likely to be obscured (Hall, Snell, & Foust, 1999); and (c) models are less likely to contain specification errors (Mulaik, 2009; Plummer, 2000). Item-level methods also have noteworthy disadvantages: (a) items are typically less reliable than are item aggregates (Kishton & Widaman, 1994); (b) using item-level data increases the rank of the covariance matrix, thereby increasing the likelihood of a poor fit (Bandalos, 2002; Bandalos & Finney, 2001); (c) item-level data are rarely multivariate normal, an important assumption underlying maximum-likelihood estimation, the most popular estimation method (Hu & Bentler, 1998).

A fourth approach is parceling (Cattell, 1974; Little et al., 2002) or partial disaggregation (Williams & O'Boyle, 2008). Parceling involves summing or averaging subsets of indicators to form aggregates larger than one item but smaller than the entire scale. Helpful guidelines exist for the use of parcels under a variety of circumstances (Coffman & MacCallum, 2005; Kishton & Widaman, 1994; Little, Lindenberger, & Nesselroade, 1999; Williams & O'Boyle, 2008). Purported advantages of parcels include improved model fit, increased communalities, enhanced reliability, normalization of indicators, smaller required sample size, and parsimony. Disadvantages include the obscuring of more appropriate measurement models, and potentially serious variability in results depending upon the allocation of items into parcels (Sterba & MacCallum, 2010). Even under optimal conditions, parceling is still only a post hoc way to obtain multiple indicators. Like all the model-based reliability correction methods, and assuming the original scale is unidimensional and devoid of method effects, parceling generates an error-corrected version of whatever the original scale measured: probably the construct of interest plus whatever other sources of variance permeate the measure.

Both of the last two approaches involve the extraction of true latent variables from multiple indicators. In latent variable structural equation modeling, one can distinguish goodness-of-fit due to the measurement model from goodness-of-fit due to the structural

[7] We thank Scott E. Maxwell for pointing this out to us.

[8] When a single indicator is used for a latent variable, only one measurement parameter may be estimated, and the others must be constrained to some value to achieve local identification. The four relevant parameters are the loading on the common factor, the common factor's variance, the loading on the unique factor, and the unique factor's variance. The loading on the unique factor is always fixed to 1. If the unique variance is fixed to a value based on reliability, we are faced with a choice—either estimate the common factor's variance and fix the loading (typically to 1) or estimate the loading and fix the common factor's variance (typically to 1). The resulting models are equivalent, so the choice is arbitrary.

model (O'Boyle & Williams, 2011a, 2011b). This represents a real advantage relative to the manifest variable path analytic methods described in the current article, in which these two features are inextricably conflated.

Test simpler models. In general, error-reduction strategies during research design and data collection are superior to error-correction strategies during data analysis. Unfortunately, most error reduction strategies require more time and money. As models become increasingly complex, these resources are often spread more thinly across the assessment of a larger number of variables, increasing the likelihood that measurement error will pervade the model. Fearing the underestimation of key theoretical paths, researchers may place greater emphasis on the measurement of variables they deem to be more important. Such practices will lead to parameter over- and underestimation, not always in the anticipated direction. Consequently, we recommend conducting simpler studies (at least when the theoretical context allows it), and spending the extra resources on the reduction of measurement error. At least three general principles favor a focus on simple models: (a) although simple models are still susceptible to measurement error, these effects often are more easily recognized and corrected in simple models than in more complex models; (b) simpler models are easier to specify, estimate, and interpret; and (c) as simple laws tend to operate under a wider variety of circumstances, more parsimonious models are likely to be more replicable and generalizable (Myung, Pitt, & Kim, 2004; Pitt & Myung, 2002). We recognize, however, that simple models may not always be compatible with certain research goals (Roberts & Pashler, 2000).

Conclusions

Uncorrected manifest variable path analyses remain commonplace in premier psychological journals despite numerous methodological articles describing the inherent dangers. We document at least five major problems. First, measurement error will almost always lead to the underestimation of some path coefficients and overestimation of other path coefficients. Second, when moderate measurement error pervades a model, spurious indications of model misfit become likely even for models that are perfectly valid when variables are measured without error. This model misfit will appear to be due to the absence of structural paths (not to the presence of measurement error), potentially leading to the addition of paths that would have been completely unnecessary had the variables been perfectly reliable. Third, when substantial measurement error pervades a model, finding evidence of model misfit becomes highly unlikely, even when the model is misspecified. Fourth, differential measurement error in various parts of a model can radically change the conclusions derived from the model. Fifth, many of these problems become larger and less anticipatable as models become more complex.

Several of these problems are worse (or at least more pervasive) than we have implied, as they pertain to manifest variable statistical methods other than path analysis. In multiple regression, differential reliability among measures of correlated predictors can change the apparent relative predictive utility of one construct over another (Cohen, Cohen, West, & Aiken, 2003). In analysis of variance (ANOVA) designs, measurement error in the operationalization of factors (e.g., forming groups artificially by dichotomizing continuous variables) can diminish power to detect effects that really do exist and spuriously increase the likelihood of detecting effects that truly do not exist (Maxwell & Delaney, 1993). In partial correlation and analysis of covariance (ANCOVA) designs, measurement error in the control variable is also a major concern (Keppel, 1991; Maxwell & Delaney, 2004; Vargha, Rudas, Delaney, & Maxwell, 1996). Although these points have been made previously, these methods continue to be used with sometimes casual regard for measurement error. How often these methods are applied with fallible measures, the magnitude of the resultant problems, and the generalizability of our recommendations to these other methodologies all represent important avenues for continued research.

The first and foremost solution to these problems is to implement more reliable measurement strategies in studies that utilize manifest variable path analysis. A second approach is to engage in any of several model-based error correction methods. Third is to obtain multiple measures and utilize latent variable data analytic methods. Ideally, both the error-reduction strategy and latent variable methods can be implemented, as recent research has revealed that measurement error can actually affect the precision of latent variable approaches (e.g., Hancock & Mueller, 2011; Heene et al., 2011; Ledgerwood & Shrout, 2011). Sometimes error reduction and latent variable strategies are not possible, in which case our fourth recommendation is to implement error correction methods, test much simpler models, and acknowledge that both of these methods have their own potentially serious liabilities.

References

Adcock, R. J. (1878). A problem in least squares. *The Analyst (Annals of Mathematics), 5,* 53–54. doi:10.2307/2635758

Baguley, T. (2009). Standardized or simple effect size: What should be reported? *British Journal of Psychology, 100,* 603–617. doi:10.1348/000712608X377117

Bandalos, D. L. (2002). The effects of item parceling on goodness-of-fit and parameter estimate bias in structural equation modeling. *Structural Equation Modeling, 9,* 78–102. doi:10.1207/S15328007SEM0901_5

Bandalos, D. L., & Finney, S. J. (2001). Item parceling issues in structural equation modeling. In G. A. Marcoulides & R. E. Schumacker (Eds.), *New developments and techniques in structural equation modeling* (pp. 269–296). Mahwah, NJ: Erlbaum.

Bedeian, A. G., Day, D. V., & Kelloway, E. K. (1997). Models: Correcting for measurement error attenuation in structural equation models: Some important reminders. *Educational and Psychological Measurement, 57,* 785–799. doi:10.1177/0013164497057005004

Bentler, P. M., & Speckart, G. (1981). Attitudes "cause" behaviors: A structural equation analysis. *Journal of Personality and Social Psychology, 40,* 226–238. doi:10.1037/0022-3514.40.2.226

Blalock, H. M., Jr. (1964). *Causal inferences in nonexperimental research*. Chapel Hill, NC: The University of North Carolina Press.

Blalock, H. M., Jr. (1965). Some implications of random measurement error for causal inferences. *American Journal of Sociology, 71,* 37–47. doi:10.1086/223991

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

Brown, W. (1910). Some experimental results in the correlation of mental abilities. *British Journal of Psychology, 3,* 296–322.

Cattell, R. B. (1974). Radial parcel factoring vs. item factoring in defining personality structure in questionnaires: Theory and experimental checks. *Australian Journal of Psychology, 26,* 103–119. doi:10.1080/00049537408255223

Coffman, D. L., & MacCallum, R. C. (2005). Using parcels to convert path analysis models into latent variable models. *Multivariate Behavioral Research, 40,* 235–259. doi:10.1207/s15327906mbr4002_4

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Erlbaum.

Cole, D. A., Ciesla, J., & Steiger, J. H. (2007). The insidious effects of completely justifiable correlated residuals in latent variable covariance structure analysis. *Psychological Methods, 12,* 381–398. doi:10.1037/1082-989X.12.4.381

Cole, D. A., & Maxwell, S. E. (2003). Testing mediational models with longitudinal data: Myths and tips in the use of structural equation modeling. *Journal of Abnormal Psychology, 112,* 558–577. doi:10.1037/0021-843X.112.4.558

Cudeck, R., & Browne, M. W. (1992). Constructing a covariance matrix that yields a specified minimizer and a specified minimum discrepancy function value. *Psychometrika, 57,* 357–369. doi:10.1007/BF02295424

Deming, W. E. (1943). *Statistical adjustment of data*. New York, NY: Wiley.

DeShon, R. P. (1998). A cautionary note on measurement error corrections in structural equation models. *Psychological Methods, 3,* 412–423. doi:10.1037/1082-989X.3.4.412

Duncan, O. D. (1975). *Introduction to structural equation models*. New York, NY: Academic Press.

Greenland, S., Schlesselman, J. J., & Criqui, M. H. (1986). The fallacy of employing standardized regression coefficients and correlations as measures of effect. *American Journal of Epidemiology, 123,* 203–208.

Hall, R. J., Snell, A. F., & Foust, M. S. (1999). Item parceling strategies in SEM: Investigating the subtle effects of unmodeled secondary constructs. *Organizational Research Methods, 2,* 233–256. doi:10.1177/109442819923002

Hancock, G. R., & Mueller, R. O. (2011). The reliability paradox in assessing structural relations within covariance structure models. *Educational and Psychological Measurement, 71,* 306–324. doi:10.1177/0013164410384856

Hayduk, L. A. (1987). *Structural equation modeling with LISREL: Essentials and advances*. Baltimore, MD: Johns Hopkins University Press.

Heene, M., Franzens, K., Hilbert, S., Draxler, C., Ziegler, M., & Bühner, M. (2011). Masking misfit in confirmatory factor analysis by increasing unique variances: A cautionary note on the usefulness of cutoff values of fit indices. *Psychological Methods, 16,* 319–336. doi:10.1037/a0024917

Hoyle, R. H., & Kenny, D. A. (1999). Sample size, reliability, and tests of statistical mediation. In R. Hoyle (Ed.), *Statistical strategies for small sample research* (pp. 195–222). Thousand Oaks, CA: Sage.

Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods, 3,* 424–453. doi:10.1037/1082-989X.3.4.424

James, L. R., Mulaik, S. A., & Brett, J. M. (1982). *Causal analysis: Assumptions, models, and data*. Beverly Hills, CA: Sage.

Kano, Y. (2007). Selection of manifest variables. In S.-Y. Lee (Ed.), *Handbook of computing and statistics with applications. Vol. 1: Handbook of latent variables and related models* (pp. 65–86). Amsterdam, the Netherlands: Elsevier.

Kenny, D. A. (1979). *Correlation and causality*. New York, NY: Wiley.

Keppel, G. (1991). *Design and analysis: A researcher's handbook* (3rd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Kim, J.-O., & Ferree, G. D., Jr. (1981). Standardization in causal analysis. *Sociological Methods & Research, 10,* 187–210.

King, D. W., King, L. A., & Foy, D. W. (1996). Prewar factors in combat-related posttraumatic stress disorder: Structural equation modeling with a national sample of female and male Vietnam veterans. *Journal of Consulting and Clinical Psychology, 64,* 520–531. doi:10.1037/0022-006X.64.3.520

King, G. (1986). How not to lie with statistics: Avoiding common mistakes in quantitative political science. *American Journal of Political Science, 30,* 666–687. doi:10.2307/2111095

Kishton, J. M., & Widaman, K. F. (1994). Unidimensional versus domain representative parceling of questionnaire items: An empirical example. *Educational and Psychological Measurement, 54,* 757–765. doi:10.1177/0013164494054003022

Koopmans, T. C. (1937). *Linear regression analysis of economic time series*. Haarlem, the Netherlands: DeErven F. Bohn.

Kummell, C. H. (1879). Reduction of observation equations which contain more than one observed quantity. *The Analyst (Annals of Mathematics), 6,* 97–105. doi:10.2307/2635646

Ledgerwood, A., & Shrout, P. E. (2011). The trade-off between accuracy and precision in latent variable models of mediation processes. *Journal of Personality and Social Psychology, 101,* 1174–1188. doi:10.1037/a0024776

Little, T. D., Cunningham, W. A., Shahar, G., & Widaman, K. F. (2002). To parcel or not to parcel: Exploring the question, weighing the merits. *Structural Equation Modeling, 9,* 151–173. doi:10.1207/S15328007SEM0902_1

Little, T. D., Lindenberger, U., & Nesselroade, J. R. (1999). On selecting indicators for multivariate measurement and modeling with latent variables: When "good" indicators are bad and "bad" indicators are good. *Psychological Methods, 4,* 192–211. doi:10.1037/1082-989X.4.2.192

Lomax, R. G. (1986). The effect of measurement error in structural equation modeling. *Journal of Experimental Education, 54,* 157–162.

MacCallum, R. C. (2003). Working with imperfect models. *Multivariate Behavioral Research, 38,* 113–139. doi:10.1207/S15327906MBR3801_5

Marsh, H. W., Hau, K.-T., Balla, J. R., & Grayson, D. (1998). Is more ever too much? The number of indicators per factor in confirmatory factor analysis. *Multivariate Behavioral Research, 33,* 181–220. doi:10.1207/s15327906mbr3302_1

Maxwell, S. E., & Delaney, H. D. (1993). Bivariate median splits and spurious statistical significance. *Psychological Bulletin, 113,* 181–190. doi:10.1037/0033-2909.113.1.181

Maxwell, S. E., & Delaney, H. D. (2004). *Designing experiments and analyzing data: A model comparison perspective* (2nd ed.). Mahwah, NJ: Erlbaum.

McCallum, B. T. (1972). Relative asymptotic bias from errors of omission and measurement. *Econometrica, 40,* 757–758. doi:10.2307/1912970

Mulaik, S. A. (2009). *Linear causal modeling with structural equations*. Boca Raton, FL: CRC Press. doi:10.1201/9781439800393

Myung, I. J., Pitt, M. A., & Kim, W. (2004). Model evaluation, testing and selection. In K. Lambert & R. Goldstone (Eds.), *Handbook of cognition* (pp. 422–436). Thousand Oaks, CA: Sage.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory*. New York, NY: McGraw-Hill.

O'Boyle, E. H., & Williams, L. J. (2011a). Decomposing model fit: Measurement vs. theory in organizational research using latent variables. *Journal of Applied Psychology, 96,* 1–12. doi:10.1037/a0020539

O'Boyle, E. H., & Williams, L. J. (2011b). Correction to O'Boyle and Williams (2011). *Journal of Applied Psychology, 96,* 729. doi:10.1037/a0024273

Pitt, M. A., & Myung, I. J. (2002). When a good fit can be bad. *Trends in Cognitive Sciences, 6,* 421–425. doi:10.1016/S1364-6613(02)01964-2

Plummer, B. A. (2000). *To parcel or not to parcel: The effects of item parceling in confirmatory factor analysis* (Unpublished doctoral dissertation). University of Rhode Island, Kingston, RI.

Rigdon, E. E. (1994). Demonstrating the effects of unmodeled random measurement error. *Structural Equation Modeling, 1,* 375–380. doi:10.1080/10705519409539986

Roberts, S., & Pashler, H. (2000). How persuasive is a good fit? A comment on theory testing. *Psychological Review, 107,* 358–367. doi:10.1037/0033-295X.107.2.358

Rock, D. A., Werts, C. E., Linn, R. L., & Jöreskog, K. G. (1977). A maximum likelihood solution to the errors in variables and errors in equations models. *Multivariate Behavioral Research, 12,* 187–197. doi:10.1207/s15327906mbr1202_6

Rubio, D. M., & Gillespie, G. F. (1995). Problems with error in structural equation models. *Structural equation modeling, 2,* 367–378. doi:10.1080/10705519509540020

Saris, W. E., & Aalberts, C. (2003). Different explanations for correlated disturbance terms in MTMM studies. *Structural Equation Modeling, 10,* 193–221. doi:10.1207/S15328007SEM1002_2

Shevlin, M., & Miles, J. N. V. (1998). Effects of sample size, model specification and factor loadings on the GFI in confirmatory factor analysis. *Personality and Individual Differences, 25,* 85–90. doi:10.1016/S0191-8869(98)00055-5

South, S. C., Krueger, R. F., & Iacono, W. G. (2011). Understanding general and specific connections between psychopathology and marital distress: A model based approach. *Journal of Abnormal Psychology, 120,* 935–947. doi:10.1037/a0025417

Spearman, C. (1910). Correlation calculated from faulty data. *British Journal of Psychology, 3,* 271–295.

Stephenson, M. T., & Holbert, R. L. (2003). A Monte Carlo simulation of observable- versus latent-variable structural equation modeling techniques. *Communication Research, 30,* 332–354. doi:10.1177/0093650203030003004

Sterba, S. K., & MacCallum, R. C. (2010). Variability in parameter estimates and model fit across repeated allocations of items to parcels. *Multivariate Behavioral Research, 45,* 322–358. doi:10.1080/00273171003680302

Taasoobshirazi, G., & Carr, M. (2009). A structural equation model of expertise in college physics. *Journal of Educational Psychology, 101,* 630–643. doi:10.1037/a0014557

Tomarken, A. J., & Waller, N. G. (2003). Potential problems with "well-fitting" models. *Journal of Abnormal Psychology, 112,* 578–598. doi:10.1037/0021-843X.112.4.578

Tucker, L. R, Koopman, R. F., & Linn, R. L. (1969). Evaluation of factor analytic research procedures by means of simulated correlation matrices. *Psychometrika, 34,* 421–459. doi:10.1007/BF02290601

Vargha, A., Rudas, T., Delaney, H. D., & Maxwell, S. E. (1996). Dichotomization, partial correlation, and conditional independence. *Journal of Educational and Behavioral Statistics, 21,* 264–282.

Wansbeek, T., & Meijer, E. (2000). *Measurement error and latent variables in econometrics*. Amsterdam, the Netherlands: Elsevier.

Werts, C. E., & Linn, R. L. (1970). Path analysis: Psychological examples. *Psychological Bulletin, 74,* 193–212. doi:10.1037/h0029778

Wickens, M. R. (1972). A note on the use of proxy variables. *Econometrica, 40,* 759–761. doi:10.2307/1912971

Williams, L. J., & O'Boyle, E. H., Jr. (2008). Measurement models for linking latent variables and indicators: A review of human resource management research using parcels. *Human Resource Management Review, 18,* 233–242. doi:10.1016/j.hrmr.2008.07.002

Wolfle, L. M. (1979). Unmeasured variables in path analysis. *Multiple linear regression viewpoints, 9,* 20–56.

Wolfle, L. M. (1980). Unmeasured variables in path analysis: Addendum. *Multiple Linear Regression Viewpoints, 10,* 61–63.

Received August 4, 2012
Revision received May 13, 2013
Accepted May 25, 2013 ■

Correction to Cole and Preacher (2013)

The article, “Manifest Variable Path Analysis: Potentially Serious and Misleading Consequences Due to Uncorrected Measurement Error” by David A. Cole and Kristopher J. Preacher (*Psychological Methods*, Advanced online publication, September 30, 2013. doi: 10.1037/a0033805), contained several errors:

Footnote 2 should have stated, “Throughout, we assume that all path coefficients are non-negative. In this context, the words attenuation and underestimation refer to the shrinkage toward zero of path coefficient estimates due to measurement error. The words inflation and overestimation refer to the expansion away from zero of path coefficient estimates.”

Footnote 4 should have stated, “We make several assumptions in our path diagrams . . .”

The first sentence in the first full paragraph on page 5 should have stated, “Fourth, the *overestimation* of any directional path is due to the underestimation of the sum of all valid tracings responsible for the covariation between the two variables of interest (other than the tracing that consists of the target path).

All versions have been corrected.

http://dx.doi.org/10.1037/a0037174