

# The Underappreciated Effects of Unreliability on Multiple Regression and Mediation

David Trafimow

Correspondence: David Trafimow, Department of Psychology, MSC 3452, New Mexico State University, P. O. Box 30001, Las Cruces, NM 88003-8001.

Received: June 25, 2021

Accepted: July 6, 2021

Available online: July 13, 2021

doi:10.11114/afa.v7i2.5292

URL: <https://doi.org/10.11114/afa.v7i2.5292>

## Abstract

There is an increasing trend for researchers in the social sciences to draw causal conclusions from correlational data. Even researchers who use relatively causally neutral language in describing their findings, imply causation by including diagrams with arrows. Moreover, they typically make recommendations for intervention or other applications in their discussion sections, that would make no sense without an implicit assumption that the findings really do indicate causal pathways. The present manuscript commences with the generous assumption that regression-based procedures extract causation out of correlational data, with an exploration of the surprising effects of unreliability on causal conclusions. After discussing the pros and cons of correcting for unreliability, the generous assumption is questioned too. The conclusion is that researchers should be more cautious in interpreting findings based on correlational research paradigms.

**Keywords:** classical true score theory, attenuation, multiple regression, beta weights, path coefficients

## 1. Introduction

An important advance, in the early history of psychology, was the establishment of the relationship between reliability and validity, for which Charles Spearman (1904) was largely responsible. Spearman’s reasoning provided the foundation for what is now considered the classical theory of measurement (see Gulliksen, 1987; Lord & Novick, 1968; for well-cited reviews). Although the theory has been improved with the addition of stronger assumptions in more recently developed measurement theories, thereby allowing stronger conclusions, the classical measurement theory remains useful, particularly in those cases where the stronger assumptions of more modern theories are not necessary to derive the conclusions of interest. In fact, the weaker assumptions of the classical theory can be considered an advantage, as they are less likely than stronger assumptions to be false (Hulin, Drasgow, & Parsons, 1983; Lord & Novick, 1968; Trafimow, 2017). The present work is a case in point, where the classical theory is sufficient to demonstrate important potential pitfalls that unreliability poses for the interpretation of complex regression analyses.

As will become clear, unreliability has predictable effects on bivariate correlations: unreliability attenuates them (Spearman, 1904). In contrast, unreliability has unpredictable effects in multiple regression or path analytic contexts (e.g., Schmidt, Le, & Oh, 2013). In these more complex contexts, unreliability can increase or decrease beta weights or path coefficients, make them appear or disappear, or even change their sign. And yet, substantive researchers seem unaware of these interpretive problems. By exploiting the simplicity of the classical theory, the present hope and expectation is that researchers will find the demonstrations to be presented convincing, easy to understand, and of obvious relevance to substantive regression work.

## Charles Spearman

Charles Spearman (1904) provided the original proof that unreliability attenuates correlation coefficients and his work was foundational for what eventually became known as classical true score theory or classical test theory or simply the classical theory (see Gulliksen, 1987; Lord & Novick, 1968 for well-cited reviews). Equation 1 provides a convenient form of Spearman’s famous attenuation formula where the goal is to evaluate how the correlation between scores uncontaminated by random error, also termed the true correlation and symbolized as  $r_{T_x T_z}$ , work with the reliabilities of the two variables  $r_{xx'}$  and  $r_{zz'}$ , to determine the observed correlation  $r_{xz}$ :

$$r_{xz} = r_{T_x T_z} \sqrt{r_{xx'} r_{zz'}}. \tag{1}$$

A quick way to understand Equation 1 is to imagine two extremes. The first extreme is that both reliabilities are perfect, equaling 1.0, in which case the observed correlation equals the true correlation. This is the ideal case. The opposite extreme is to imagine that one of the reliabilities equals 0, in which case the observed correlation equals 0 no matter the value of the true correlation. Between the two extremes, intermediate reliability values cause the observed correlation to be attenuated relative to the true correlation. For example, if the product of the reliabilities equals 0.49, then the observed correlation will only be 70% of the true correlation. The effect of unreliability is consistent: unreliability always attenuates observed correlations relative to true correlations and more unreliability forces more attenuation.

But few researchers concern themselves with bivariate correlation coefficients. Rather, they perform more complex multiple regression analyses and attend carefully to the resulting multiple correlations or beta weights. Or, they perform mediation analyses with path coefficients, presumed to indicate causality. What are the effects of unreliability in these more complex contexts?

To commence with an immediate suspicion, consider that to determine beta weights, the researcher sets up a series of equations, where the number of equations equals the number of unknown beta-weights. By solving the simultaneous equations, the researcher can obtain the beta-weights. And there is good news and bad news. The good news is that procedures for solving simultaneous equations have existed for centuries and can be applied easily to multiple regression contexts, especially with the aid of mathematical or statistical programs. The bad news is that each value obtained depends on the other obtained values. Consequently, if there is random measurement error, there is the possibility for the obtained findings to be misleading. The present goal is to explore the interaction of the necessity that all beta weights depend on each other and the limitation that measures of variables are associated with random error (unreliability). This interaction implies the possibility of inconsistent unreliability effects, to be explored in subsequent sections.

**Standard Multiple Regression**

Consider a standard multiple regression equation where there is a criterion variable  $y$  and predictor variables 1 and 2. Based on three bivariate correlations—the correlation of each predictor variable with the criterion variable ( $r_{y1}$  and  $r_{y2}$ ) and the correlation between the predictor variables ( $r_{12}$ )—Equation 2 shows how to obtain the multiple correlation predicting the criterion variable ( $R_{y.12}$ ) from the two predictor variables (Pedhazur, 1997):

$$R_{y.12} = \sqrt{\frac{r_{y1}^2 + r_{y2}^2 - 2r_{y1}r_{y2}r_{12}}{1 - r_{12}^2}} \tag{2}$$

Pedhazur (1997) also presented straightforward equations for obtaining beta-weights for the two predictor variables and these are symbolized as  $\beta_1$  and  $\beta_2$  and obtained via Equation 3:

$$\beta_1 = \frac{r_{y1} - r_{y2}r_{12}}{1 - r_{12}^2} \text{ and } \beta_2 = \frac{r_{y2} - r_{y1}r_{12}}{1 - r_{12}^2} \tag{3}$$

To move in the direction of our goal, it is necessary to modify Equations 2 and 3 so that they account for the reliabilities with which the criterion variable and the predictor variables are measured. It is useful to apply Equation 1 to the three correlation coefficients to obtain the bullet-listed components below:

- $r_{y1} = r_{T_y T_1} \sqrt{r_{yy'} r_{11'}}$ ,
- $r_{y2} = r_{T_y T_2} \sqrt{r_{yy'} r_{22'}}$ ,
- $r_{12} = r_{T_1 T_2} \sqrt{r_{11'} r_{22'}}$ .

The bullet-listed components express the bivariate correlations used in Equations 2 and 3 in terms of true correlations and reliabilities of measures of variables. Instantiating the bullet-listed components into Equation 2 and Equation 3 render Equation 4 and Equation 5, respectively, in terms of true correlations and reliabilities as opposed to observed correlations.

$$R_{y.12} = \sqrt{\frac{r_{T_y T_1}^2 r_{yy'} r_{11'} + r_{T_y T_2}^2 r_{yy'} r_{22'} - 2r_{T_y T_1} \sqrt{r_{yy'} r_{11'}} r_{T_y T_2} \sqrt{r_{yy'} r_{22'}} \sqrt{r_{11'} r_{22'}}}{1 - r_{T_1 T_2}^2 r_{11'} r_{22'}} \tag{4}$$

$$\beta_1 = \frac{r_{T_y T_1} \sqrt{r_{yy'} r_{11'}} - (r_{T_y T_2} \sqrt{r_{yy'} r_{22'}}) (r_{T_1 T_2} \sqrt{r_{11'} r_{22'}})}{1 - r_{T_1 T_2}^2 r_{11'} r_{22'}} \tag{5a}$$

$$\beta_2 = \frac{r_{T_y T_2} \sqrt{r_{yy'} r_{22'}} - (r_{T_y T_1} \sqrt{r_{yy'} r_{11'}}) (r_{T_1 T_2} \sqrt{r_{11'} r_{22'}})}{1 - r_{T_1 T_2}^2 r_{11'} r_{22'}} \tag{5b}$$

## Multiple Correlations

Let us commence with investigating the effects of unreliability on multiple correlations, using Equation 4. The effects of unreliability on beta weights will be addressed in the subsequent subsection.

Figure 1 illustrates a case where all true correlations are set at 0.5, and the reliability of either the criterion variable or one of the predictor variables allowed to vary between 0 and 1, keeping the reliability of the other variable perfect. Although, in the case of a bivariate correlation, the reliabilities of the predictor and criterion variables are equally important, this is not true for a multiple correlation. Figure 1 shows clearly that varying the reliability of one of the predictor variables, keeping the reliability of the other predictor variable and the criterion variable at 1, only slightly influences the multiple correlation. That is, the multiple correlation ranges from 0.500 when the reliability of one of the predictor variables is 0, to 0.577 when it is 1. In contrast, keeping the reliabilities of both predictor variables at 1, the multiple correlation ranges from 0 when the reliability of the criterion variable is 0, to 0.577 when it is 1. Thus, we already see that unreliability can work very differently for multiple correlations than for bivariate ones.

In contrast to Figure 1, where all true correlations were set equal to each other at the 0.5 level, they are very unequal in Figure 2. Specifically, the true correlation between the first predictor variable and the criterion variable is set at 0.8, the true correlation between the second predictor variable and the criterion variable is set at 0.2, and the true correlation between the predictor variables is set at 0.7. The top curve in Figure 2 depicts that unreliability in the second predictor only has a small effect on the multiple correlation, letting the other two variables remain perfectly reliable. The reason is that the second predictor variable is only a poor predictor of the criterion, even with perfect reliability, so unreliability has only a small effect on the multiple correlation. In contrast, the first predictor variable is an excellent predictor of the criterion variable, so introducing unreliability matters much, as it also does for the criterion variable. It is interesting that whether unreliability matters more when it applies to the first predictor variable or the criterion variable depends on the range of consideration, as Figure 2 shows crossing curves when reliability is low. Both Figure 1 and Figure 2 show that the effects of unreliability are less consistent when they pertain to multiple correlations than to bivariate correlations. Nevertheless, there is one consistency even so, which is that unreliability usually decreases, and never increases, both multiple correlations and bivariate correlations. Matters become more contradictory when we consider beta weights in the subsequent subsection.

## Beta Weights

Let us commence with a case similar to that used in Figure 1, where all true correlations are set at 0.5. Figure 3 shows how varying the reliability of the second predictor variable, keeping the other two reliabilities at unity, influences the first beta weight (Beta 1) and the second beta weight (Beta 2). Note the different conclusions that unreliability implies. For example, when all reliabilities are perfect, thereby reflecting the implications of the true correlations, both beta weights equal each other. However, decreasing the reliability of the second predictor variable simultaneously causes its beta weight to decrease, and the beta weight of the other predictor variable to increase. Thus, unreliability in one of the predictor variables implies two errors: too small a beta weight for one predictor variable and too large a beta weight for the other.

Figure 4 shows a beta weight crossover effect, whereby flip-flopping the reliabilities of the two variables causes their respective beta weights to change. In Figure 4, all true correlations were set at 0.7. The reliability of the first predictor variable was set at 0.6, 0.7, 0.8, 0.9, or 1.0 whereas the reliability of the second predictor variable was set at 1.0, 0.9, 0.8, 0.7 or 0.6. Thus, the difference was -0.4, -0.2, 0, 0.2, or 0.4. Figure 4 shows that as the magnitude of either the positive or negative difference increases, the beta weights diverge.

Figures 5 and 6 are based on uneven true correlations, where the true correlation between the first predictor variable and the criterion variable is 0.8, the true correlation between the second predictor variable and the criterion variable is 0.2, and the true correlation between the two predictor variables is 0.7. The difference between the two figures is that in Figure 5, the reliability of the second predictor variable was allowed to deviate from perfection; whereas, in Figure 6, it was the reliability of the first predictor variable that was allowed to deviate from perfection. Recall that in Figure 3, increasing the reliability of the second predictor variable decreased the beta weight of the first predictor variable. In Figure 5, the opposite is true; increasing the reliability of the second predictor variable increases the beta-weight of the first predictor variable! The effects of unreliability on beta weights are indeed difficult to predict! And the drama increases upon investigation of the beta weight for the second predictor variable: as the reliability of the second predictor variable increases, its beta weight becomes increasingly negative!

In Figure 6, we see an unsurprising effect whereby increasing the reliability of the first predictor increases its beta weight. But there is a surprising effect that increasing the reliability of the first predictor variable causes the beta weight of the second predictor variable to become increasingly negative. Or saying the same thing, but in the opposite direction, decreasing reliability in the first predictor variable renders the beta weight of the second predictor variable to become

less negative, even to the point of eventually crossing into positive territory, and even exceeding the beta weight of the first predictor variable!

Figures 1-6 illustrate the surprising effects of unreliability in multiple regression contexts. Put simply, these surprising results indicate that multiple correlations and especially beta weights should not be taken at face value under conditions where one or more of the variables is measured with substantially less than perfect reliability. However, not taking them at face value involves much complexity that is absent in most substantive work involving multiple regression paradigms. There is a potential implication that unreliability might render the results of standard mediation analyses problematic too, an implication that deserves its own section.

**Standard Mediation Analyses**

In many correlation-based articles, researchers wish to draw causal conclusions from correlational data. Even in the rare cases where no causal conclusions are drawn when presenting the path diagrams, researchers give the show away when they make recommendations in the discussion section that presume causal interpretation of the data. In the present section, the larger philosophical issues pertaining to causation will be ignored. That correlational data can be reasonably interpreted causally via mediation analyses will be assumed, for now. However, even under this extremely generous assumption, there remains a question: Does unreliability influence causal conclusions?

To answer this question, it is useful to (a) include standard mediation equations and (b) modify them to reflect true correlations and reliabilities as opposed to observed correlations.

Equations 6-8 provide standard path coefficient equations for mediation analyses (Pedhazur, 1997) where  $p_{21}$  represents the path coefficient for the effect of Variable 1 on Variable 2,  $p_{31}$  represents the path coefficient for the effect of Variable 1 on Variable 3, and  $p_{32}$  represents the path coefficient for the effect of Variable 2 on Variable 3:

$$p_{21} = \beta_{21} = r_{21}, \tag{6}$$

$$p_{31} = \beta_{31.2} = \frac{r_{31} - r_{32}r_{12}}{1 - r_{12}^2}, \tag{7}$$

$$p_{32} = \beta_{32.1} = \frac{r_{32} - r_{31}r_{12}}{1 - r_{12}^2}. \tag{8}$$

As in the multiple regression section, it is convenient to bullet-list components to instantiate into Equations 6-8 to obtain the desired equations. The components will be listed first, and then the equations:

- $r_{21} = r_{T_2T_1} \sqrt{r_{22'}r_{11'}}$ ,
- $r_{31} = r_{T_3T_1} \sqrt{r_{33'}r_{11'}}$ ,
- $r_{32} = r_{T_3T_2} \sqrt{r_{33'}r_{22'}}$ .

Instantiating the bullet-listed components into the proper slots in Equations 6-8 results in Equations 9-11:

$$p_{21} = r_{T_2T_1} \sqrt{r_{22'}r_{11'}}, \tag{9}$$

$$p_{31} = \frac{r_{T_3T_1} \sqrt{r_{33'}r_{11'}} - r_{T_3T_2} \sqrt{r_{33'}r_{22'}} r_{T_2T_1} \sqrt{r_{22'}r_{11'}}}{1 - r_{T_2T_1}^2 r_{22'}r_{11'}}, \tag{10}$$

$$p_{32} = \frac{r_{T_3T_2} \sqrt{r_{33'}r_{22'}} - r_{T_3T_1} \sqrt{r_{33'}r_{11'}} r_{T_2T_1} \sqrt{r_{22'}r_{11'}}}{1 - r_{T_2T_1}^2 r_{22'}r_{11'}}. \tag{11}$$

With the aid of Equations 9-11, it is now possible to construct various path diagram scenarios. An accompanying figure will illustrate the consequences of unreliability in each scenario.

**Scenario 1.**

Suppose that the true correlation between Variable 1 and Variable 2 is 0.9, the true correlation between Variable 3 and Variable 1 is 0.1, and the true correlation between Variable 2 and Variable 3 is 0.1. If we assume perfect reliabilities, then the path coefficients come out as follows:  $p_{21} = 0.900$ ,  $p_{31} = 0.053$ , and  $p_{32} = 0.053$ . However, suppose we reduce the reliabilities of Variable 1 and Variable 3 to 0.60. In that case, the path coefficients change:  $p_{21} = 0.697$ ,  $p_{31} = 0.012$ , and  $p_{32} = 0.069$ . Note that decreasing these reliabilities decreased two of the path coefficients and increased one of them. It is particularly noteworthy that the direct path from Variable 1 to Variable 3 almost completely disappeared, thereby showing that unreliability can cause path coefficients that are there to reduce to the point of disappearance. Figure 7 illustrates this effect, where the first number along each pathway represents the path coefficient that would be obtained with perfect reliabilities and the second number represents the path coefficient that would be obtained with the reliabilities in the scenario, in boldface. Note that a similar disappearing of the path from Variable 2 to Variable 3 would occur by setting the reliabilities of Variable 2 and Variable 3 at the 0.60 level, keeping the reliability of Variable 1 at unity. Thus, unreliability can force causal paths that are there to decrease or practically disappear.

## Scenario 2

Scenario 1 demonstrated how unreliability can make causal paths disappear, but this was not a particularly interesting demonstration as it is already well-known that, even in the case of a bivariate correlation, unreliability can force substantial attenuation. It would be more interesting if unreliability could be shown to make causal paths appear that otherwise are not there. This would be a dramatic departure from the bivariate case. In addition, because the typical goal of researchers who perform path analyses is to show that hypothesized causal pathways are supported by the data, it would be particularly inconvenient for such research if it could be demonstrated that unreliability creates causal pathways that are not there according to the true correlations.

But is it possible? Consider a scenario where the true correlation between Variable 1 and Variable 3 is 0.555, the true correlation between Variable 1 and Variable 2 is 0.500, and the true correlation between Variable 2 and Variable 3 is 0.90. A mediation analysis renders the following path coefficients under perfect reliability: the path coefficient from Variable 1 to Variable 2 is 0.555, the direct path coefficient from Variable 1 to Variable 3 is 0.001 (practically zero), and the path coefficient from Variable 2 to Variable 3 is 0.900. However, suppose that we let the reliability of Variable 2 drop to 0.60. In that case, the path coefficient from Variable 1 to Variable 3 rises to 0.246. In essence, unreliability has created a direct (unmediated) pathway from Variable 1 to Variable 3 out of thin air! Figure 8 illustrates the effect. Note, too, the strong decrease in the path coefficient from Variable 2 to Variable 3, from 0.900 to 0.592.

## Scenario 3

However, even Scenario 2 is not dire for many mediation researchers because the typical goal is not to demonstrate a direct causal pathway from Variable 1 to Variable 3, but rather a mediated pathway from Variable 1 to Variable 2, and from Variable 2 to Variable 3. The key issue, then, is whether it is possible to generate a causal pathway from Variable 2 to Variable 3 out of thin air.

In fact, this is possible. Consider a scenario where the true correlations involving Variable 1 are set at 0.5 and the true correlation between Variable 2 and Variable 3 is set at 0.25. With perfect reliabilities, the path coefficient from Variable 1 to Variable 2 is 0.5, the path coefficient from Variable 1 to Variable 3 is 0.5, and the path coefficient from Variable 2 to Variable 3 is 0. Thus, there is no mediation. But now, suppose that we let the reliability of Variable 1 drop to 0.6. In that case, the path coefficient from Variable 2 to Variable 3 increases from 0 to 0.118. Thus, we see that unreliability can generate mediation out of thin air. Figure 9 illustrates the path diagram. Figure 9 also shows that unreliability in Variable 1 forces pathways from it to both of the other variables to decrease.

## Scenario 4

Suppose that the true correlation between Variable 1 and Variable 2 is 0.90, the true correlation between Variable 1 and Variable 3 is 0.13, and the true correlation between Variable 2 and Variable 3 is 0.10. In that case, under perfect reliability, the path coefficient from Variable 1 to Variable 2 is 0.900, the path coefficient from Variable 1 to Variable 3 is 0.211, and the path coefficient from Variable 2 to Variable 3 is -0.089. In terms of mediation, Variable 1 is a positive cause of Variable 2, but Variable 2 is a negative cause of Variable 3. However, reducing the reliability of Variable 1 to 0.60 causes all the path coefficients to change, as Figure 10 illustrates, and the path coefficient from Variable 2 to Variable 3 changes from -0.089 to 0.058. In essence, unreliability can change the sign of the mediated pathway.

## Scenario 5

It sometimes happens that the researcher is confident that Variable 1 is the cause of Variables 2 and 3 but is unsure about whether there is mediation through Variable 2 or through Variable 3. The researcher is interested in whether  $1 \rightarrow 2 \rightarrow 3$  or  $1 \rightarrow 3 \rightarrow 2$  is a better model. The researcher would like to try out both models and choose that which has the most impressive mediated pathway.

Suppose that all true correlations equal 0.70, with perfect reliabilities. In that case, both models perform equally. The path coefficient from Variable 1 to either Variable 2 or Variable 3 is 0.700 and the other path coefficients are 0.412 as Figure 11 shows. But suppose that either Variable 2 or Variable 3 has reliability set at 0.6, how do the two models fare?

As usual, the values exhibited in Figure 11 that are in normal type represent the case when all reliabilities are perfect, as explained in the foregoing paragraph. The italicized values in Figure 11 represent when the reliability of Variable 2—the mediating variable—is set at 0.6, keeping the others perfect, with three important effects. First, the path coefficient from Variable 1 to Variable 2 decreases from 0.700 to 0.542. Second, the path coefficient from Variable 1 to Variable 3 increases from 0.412 to 0.575; as we have seen before, decreasing the reliability of Variable 2 increases the direct effect from Variable 1 to Variable 3. Third, the path coefficient from Variable 2 to Variable 3 decreases from 0.412 to 0.230. In contrast, setting the Variable 3 reliability to 0.6, keeping the others perfect, results in a much stronger mediated pathway and a weaker direct pathway (see boldfaced values in Figure 11). First, the path coefficient from Variable 1 to Variable 3 remains at 0.700. Second, the path coefficient from Variable 1 to Variable 3 decreases from 0.412 to 0.319, and this is

more of a decrease than when we had set the reliability of Variable 2 to 0.6. Third, the path coefficient from Variable 2 to Variable 3 decreases from 0.412 to 0.319, but this is less of a decrease than when we had set the reliability of Variable 2 to 0.60. Finally, it is interesting to compute the size of the mediated pathways under the condition when the reliability of Variable 2 is set at 0.6 or under the condition when the reliability of Variable 3 is set at 0.6. In the former case, the total mediated pathway is as follows: *total mediated pathway* =  $p_{21}p_{32} = 0.125$ . In the latter case, the total mediated pathway is as follows: *total mediated pathway* =  $p_{21}p_{32} = 0.223$ . Put generically, whether Variable 2 or Variable 3 performs best as the mediator depends, in important part, on how reliably they are measured.

### Dis-Attenuating Correlations

The obvious solution to the problems illustrated by Figures 1-11 is to dis-attenuate the bivariate correlations that go into multiple regression analyses. This can be performed easily based on algebraically rearranging Equation 1 to obtain Equation 12:

$$r_{T_x T_z} = \frac{r_{xz}}{\sqrt{r_{xx'} r_{zz'}}} \quad (12)$$

Equation 12 informs us that it is possible to estimate the true bivariate correlation between two variables simply by dividing the observed bivariate correlation by the square root of the product of the reliabilities. In turn, such dis-attenuated correlations can replace observed correlations in multiple regression analyses.

Although dis-attenuating observed correlations may increase accuracy (Schmidt & Hunter, 1996; 1999; Schmidt et al., 2013), the benefits are limited by important issues, including the limitations of the classical theory (Borsboom, 2005; Borsboom & Mellenbergh, 2002; 2003; Raykov & Marcoulides, 2011; Zimmerman, 1975). As stated earlier, but it bears reiterating, the classical theory is a weak theory, with minimal assumptions. The advantage of the minimal classical assumptions is that they are less likely than the assumptions of stronger theories, such as item response theory, to be false (Hulin, Drasgow, & Parsons, 1983; Lord & Novick, 1968; Trafimow, 2017). But there are two classes of disadvantages too. One of these pertains to the many tasks that more modern theories—due to their stronger assumptions—can accomplish that the classical theory cannot accomplish. For present purposes, this issue is of low relevance because the classical theory is sufficient for the foregoing demonstrations. But a second class of disadvantages is of much present relevance. Put simply, the usual reason that researchers test participants is to learn about a construct: attitude, trait, etc. But the assumptions of the classical theory are so minimal that they do not even include the assumption that an underlying construct exists, nor are true scores necessarily equivalent to construct scores (Borsboom, 2005; Lord & Novick, 1968; Schmidt et al., 2013). Thus, it is fair to ask someone who would dis-attenuate, exactly what construct she is dis-attenuating for. For example, if the goal is to predict behavioral intentions from attitudes, the researcher wishes to assume that there are underlying constructs that can be termed behavioral intentions and attitudes, respectively. Although item response theory explicitly assumes the existence of the latent construct, the classical theory does not, and it is not necessarily clear, at the theoretical end, that for which the researcher is dis-attenuating. It might be difficult to argue that one is dis-attenuating the correlation between attitude and behavioral intention if one is not even assuming these constructs exist. A potential counterargument is that the researcher may be assuming the existence of the constructs of interest, even if the classical dis-attenuation formula does not, though the fact that true scores need not reflect construct scores is potentially inconvenient for the counterargument. The degree to which the classical dis-attenuation formula maps onto the researcher's constructs is a debatable topic (Borsboom, 2005; Borsboom & Mellenbergh, 2002; 2003; Lord & Novick, 1968; Raykov & Marcoulides, 2011; Schmidt et al., 2013; Zimmerman, 1975). Then, too, this debate sometimes includes the issue of whether dis-attenuating, whether via the classical theory or a more modern one, makes use of fictional (unobserved) values that reduce scientific objectivity.

Another potential problem with dis-attenuating, whether using the classical theory or a more modern one, is that one needs data to do so. Specifically, it is necessary to collect the data upon which (a) the observed correlation can be estimated, (b) the reliability coefficient for one of the variables can be estimated, and (c) the reliability coefficient for the other of the variables can be estimated. If any of these values are missing, there can be no dis-attenuation. These necessities beg consideration of how well any of the three estimates can be trusted. For example, it is well-known that if the reliability estimates are too low, one could overcorrect the observed correlation, and it is even possible to obtain estimated true correlations in excess of unity. In that event, the researcher is faced with an unpleasant choice. The researcher can live with an estimated true correlation in excess of unity or the researcher can cap at unity, but lose out with respect to the desirable fact that, whatever the other problems are, the expectation of dis-attenuated correlations should approach the true correlation over indefinite trials.

The issue of overcorrection can be considered a specific case of a more general problem. That is, the observed correlations and reliability coefficients obtained in particular samples may be argued to be unlikely to be close to their corresponding population parameters. Accepting that argument implies that the estimated true correlations will be too high or too low. In turn, instantiating these into multiple regression equations will result in misleading findings. And

thanks to a recent advance, termed the a priori procedure (APP), it is possible to explore this issue in some detail (see Trafimow, 2019 for a review).

The APP is a pre-data procedure. The researcher specifies two items.

- Precision: the desired distance between the sample statistic and its corresponding population parameter.
- Confidence: the desired probability of meeting the precision specification.

Given that the researcher has specified precision and confidence and found the appropriate APP equation for the statistic of concern, the APP equation provides the required sample size. In addition, the APP can be used post-data in the sense that given the sample size used in the study, APP equations can be reverse engineered to render precision. In fact, Trafimow and Myüz (2019) and Trafimow, Hyman, and Kostyk (2020) used the APP, post-data, to demonstrate imprecision in published research in psychology and marketing, respectively.

Wang et al. (2021) derived and explained the mathematics behind using the APP for correlations. For present purposes, it is sufficient to use the link to a free and user-friendly computer program to find the sample sizes necessary to obtain various levels of precision at the typical 95% confidence level or a 90% confidence level that can be used for the sake of comparison.<sup>1</sup> Figure 12 presents the findings. At the typical 95% confidence level, it takes a sample size of 802 to obtain a sample observed correlation that is within one-tenth of a standard deviation of the population correlation. If we are willing to settle for obtaining a sample observed correlation that is within two-tenths of a standard deviation of the population correlation, the required sample size drops to 201. Of course, decreasing the precision requirement or decreasing the confidence requirement results in smaller required sample sizes, but with the penalty that the sample statistics cannot be trusted as good estimators of corresponding population parameters. And if the sample reliabilities poorly estimate the population reliabilities, then dis-attenuating based on sample reliabilities may constitute poor practice. The larger point that Figure 12 illustrates is that, even generously assuming that the sample is perfectly representative of population, sample observed correlations or reliabilities are likely to deviate considerably from population observed correlations or reliabilities under typical sample sizes. Therefore, even under the best of conditions, dis-attenuated correlations are likely to be at least somewhat inaccurate.

Worse yet, researchers rarely have the best of conditions. It is unlikely, in the extreme, that the researcher has sampled randomly and independently from the population, which adds a source of systematic error in addition to the problem of random error addressed by the APP. And, on top of the problems of both random and systematic errors, there is the issue that with the extensive multiplying of variables in equations pertaining to multiple correlations, beta weights, and path coefficients, even small errors can multiply out of control.

To be sure, researchers could make strong efforts to have measures that are as reliable as possible to minimize the potential problems associated with dis-attenuation. One way to accomplish this is to have much longer tests or to have more indicator variables. As reliability depends on the number of items (indicator variables) and interitem or inter-indicator correlations, having longer tests not only makes sense, but the beneficial effect has been observed in a variety of areas. Possibly the most dramatic psychology example occurred in the context of the trait-situation debate, whether behaviors are more determined by traits or situations. Although many found traits to be poor predictors of behaviors (see Mischel, 1968 for an influential review), Epstein (1979; 1980) showed that when personality researchers lengthened their tests, thereby increasing reliability, behavior prediction increased too.

That said, increasing test lengths can have disadvantages as well as advantages. Not only are there the obvious issues of participant time, fatigue, and so on; but there is an additional issue that with each added item or indicator variable, the researcher risks an amalgamation problem. That is, the added items may be measuring different, though correlated constructs. The constructs might be sufficiently correlated to enable the researcher to enjoy an increased reliability coefficient, while nevertheless decreasing construct clarity by amalgamating constructs that should be kept separate, and that are distinguishable by factor analytic procedures, or even by experiment. Work on the construct of perceived behavioral control provides a case in point, where the measure of the construct includes items focusing on the ease-difficulty of performing a behavior but also includes items focusing on the person's capability-incapability to perform it. Not only were the two constructs separable by factor analytic procedures, but Trafimow et al. (2002) performed experimental manipulations that influenced one construct, without influencing the other; or that influenced the other, without influencing the one. The combination of the factor analytic evidence and the experimental double dissociation provided a powerful case against the amalgamated construct of perceived behavioral control and in favor of two different constructs concerned with ease-difficulty and capability-incapability.

An argument in favor of amalgamation is that, in the interest of parsimony, scientists should not proliferate constructs (e.g., Schmidt et al., 2013). However, there are problems with this argument. Consider the case of chemistry, where

---

<sup>1</sup> The program can be found here: <https://appforsn.shinyapps.io/ncorr/>.

there are 118 elements in the periodic table, 91 of which can be found naturally. It would not be a favor to chemistry to reduce the number of elements in the interest of parsimony. Like it or not, there are many elements. And like it or not, the construct of ease-difficulty is different from the construct of capable-incapable. There are many constructs and reducing their number in the interest of parsimony merely impairs the accuracy with which researchers represent reality. Parsimony is no virtue in this context. In contrast, a context where parsimony might be a virtue is in the search for basic principles. For example, there are basic quantum principles underlying why the elements work the way they do, and parsimony may—or may not—be virtuous at the more basic level. Much may depend on whether one is a realist or a pragmatist. If one is a realist, the justification of parsimony would have to be that a parsimonious theory is more likely to be true than a less parsimonious one; the more components a theory contains, the more likely it is that at least one of them is wrong (e.g., Saylor & Trafimow, in press; Trafimow, 2017a). However, it is equally possible to argue that reality itself is complex, and so there is no way that a parsimonious theory can be correct (e.g., Cartwright, 1983; 1999). In this sense, parsimony does not go well with realism, and much debate might revolve around how one conceptualizes what it means for a theory to be true. From a pragmatist standpoint, parsimony is easier to defend because there is no assumption that the theory ought to be true. However, most of the researchers who argue for latent constructs also have committed themselves to realism (see Borsboom, 2005; Borsboom & Mellenbergh, 2003 for reviews). And as we have seen, from a realist perspective, the proscription against a multiplicity of constructs for the sake of parsimony is difficult to justify.<sup>2</sup>

### Causation

Until now, the present work has generously assumed that mediation analysis really does indicate causation, thereby providing the opportunity to demonstrate the many unappreciated difficulties that unreliability generates, even under an extremely generous assumption. Being less generous suggests yet more problems. For example, most theories are within-participants theories: each person's standing on the introversion-extraversion dimension causes behavior that is varying degrees of introverted-extraverted, each person's attitude towards exercising causes the amount of exercising that person performs, and so on. And yet, the causation in path analysis is between-participants. Even if there is a statistically significant path coefficient, and even if we ignore all the reliability-related issues discussed earlier, it is not clear for what proportion of the participants the hypothesized causation is really so. That researchers have found that within-participants regression analyses generally fail to replicate between-participants regression analyses (Molenaar, 2004; 2015), provides a strong empirical basis for doubting that path analytic causation has anything to do with within-participants causation at the theoretical level. Moreover, Grice (2011; Grice et al., 2021) has demonstrated a lack of matching of the causal implications of summary statistics (e.g., beta weights, path coefficients, etc.) and the frequencies of participants whose data patterns are consistent with them.

Another complication is that Woodside (2013a; 2013b; 2015) has argued that most effects have a multiplicity of causes that combine in a configural manner, not a dimensional manner. Woodside's argument is consistent with philosophers, who also have argued for configural thinking with respect to causation. Mackie's (1974) INUS specification may be the most famous philosophical argument in this direction, but there are many (Beebe, Hitchcock, & Menzies, 2009). Such configural thinking is not well captured by standard path analytic procedures, and alternative statistical methods have been developed to address configural causation (Woodside, 2013a; 2013b; 2015).

Finally, there is the proverbial elephant-in-the-room, that which Spirtes, Glymour, and Scheines (2000) termed the problem of *statistical indistinguishability*. To quote, "Without experimental manipulations, the resolving power of any possible method for inferring causal structure from statistical relationships is limited by statistical indistinguishability. If two causal structures can equally account for the same statistics, then no statistics can distinguish them" (p. 59). And there have been many demonstrations that so-called causal analyses are susceptible to multiple causal interpretations (e.g., Fiedler, Schott, & Meiser, 2011; Grice, Cohn, Ramsey, & Chaney, 2015; Kline, 2015; MacKinnon, Krull, & Lockwood, 2000; Tate, 2015; Thoemmes, 2015; Trafimow, 2015). Thus, the cliché that zero-order correlations fail to demonstrate causation applies to complex correlational analyses too, in the absence of experimental manipulations.

As a dramatic example of how mediation analyses can go horribly wrong, even under near-perfect reliability, Trafimow (2015) performed two mediation analyses with respect to the planets of the solar system, to test two blatantly wrong mediation models. First, he tested that planetary mass causes planetary velocity which, in turn, causes planetary momentum. Second, he tested that planetary velocity causes planetary mass which, in turn, causes planetary momentum. The mediation analyses provided strong support for the latter over the former model. But as both models are incorrect, we have a demonstration not only of the general problem of statistical indistinguishability (Spirtes et al., 2020); but of a

---

<sup>2</sup> Most modern measurement theories assume a latent construct, often symbolized as  $\theta$ . However, if one is not a realist, then there is no reason to assume that the latent construct  $\theta$  exists. And in that case, the researcher is thrown back to the classical theory.



special case of that problem too, that even competitive model tests can lead to dramatically wrong conclusions.

### **Dichotomania**

And then there is the unavoidable issue of dichotomania on the part of many researchers who use mediation analyses. Kline (2015) emphasized the problems with using statistical significance, or lack thereof, to draw dichotomous decisions about path coefficients. From an applied point of view, this is obvious anyhow. If one is going to assume that the obtained path coefficient really does give the strength of the causal relationship, there is the crucial issue of whether it is strong enough to justify an application. Once this is admitted, the problems rendered in the figures, where unreliability can have impressive impacts on the sizes (or even directions) of obtained path coefficients, are devastating for interpretation.

But perhaps dichotomania can be justified from a theoretical perspective. That is, from a theoretical perspective, a researcher might simply be concerned with whether a path is there or is not there. The issue, say, is not how much mediation there is but rather whether there is mediation at all. In turn, if the significance test supports the theorized mediation, that is all that is necessary.

But the argument is strongly problematical. To see why, consider the famous Michelson and Morley (1887) experiment that disconfirmed the existence of the luminiferous ether theorized to be necessary for the propagation of light waves. Although significance testing was not in vogue in 1887, researchers agreed that the findings were sufficiently close to null that they could be considered null. Today, most physicists agree that there is no luminiferous ether and so the theoretical effect size really is zero. However, the effect Michelson and Morley obtained with their interferometer was not null. Carver (1993) reanalyzed the thousands of data points using the modern null hypothesis significance testing procedure and obtained statistical significance. One imperfection in the experiment was a lack of control for temperature change in the device, which accounts for the slightly non-null result. Therefore, although it is easy to say that anything non-null demonstrates the theory, the history of physics indicates otherwise. Because there are always imperfections in the procedure or measurements, a statistically significant effect does not necessarily indicate the existence of a theorized entity or theorized causal path. Very small effects can easily be explained away as due to study imperfections, rather than because the theorized entity is really there. As the Michelson and Morley (1887) case illustrates, the mere fact of statistical significance fails to provide impressive support that an effect is really there: empirical imperfections could be responsible. Moreover, consistent with the theme of study imperfections, the foregoing demonstrations illustrate that unreliability can create mediation out of thin air.

### **2. Conclusion**

Over the past few decades, fancy multiple regression approaches, mediation analyses, and so on have become de rigueur. Even under the too-generous assumption that correlation really does indicate causation in mediation-type paradigms, we have seen that unreliability, which is generally present, plays havoc with interpretation. To be fair, it is possible to account for unreliability. Nevertheless, it would require unrealistic optimism to believe that available procedures that account for unreliability work even close to perfectly, and without introducing additional problems of their own. And Figure 12 illustrates that it would take many more participants than researchers typically collect, to obtain sample reliabilities that are good estimates of corresponding population reliabilities. Because accounting for unreliability necessitates assuming that sample reliabilities are good estimates of corresponding population reliabilities, this is a serious problem that has not been properly addressed in the literature.

Worse yet, the generous assumption that correlation really does indicate causation is blatantly problematic. There is the problem that most theorized causation is within-participants and yet the analyses are between-participants, the problem that many—or most—causes are configural rather than dimensional, and the ubiquitous problem of statistical indistinguishability. All these haunt typical regression and mediation approaches to data analysis.

Possibly the most typical retort to criticisms is to assert that they fail to properly credit the role of theory. Theory, according to the standard retort, justifies drawing causal conclusions from a matrix of correlations that provide the substrate for complex causal diagrams. But the retort usually fails. In the vast majority of cases, the causal diagram **is** the theory. Therefore, the retort becomes circular in that the diagram justifies itself! One reason that the causal diagram and the theory are usually the same thing is because the arrows that compose the causal diagram are usually the result of a hodgepodge of previous findings and previously asserted hypotheses, and not because of a derivation from a single umbrella theory.

An exception to the foregoing assertion might be work pertaining to the reasoned action approach (Fishbein & Ajzen, 1975; 2010). In this case, there really is an umbrella theory, and taking the theory seriously provides a strong indication that at least some of the arrows that the theory indicates are possible likely are true for particular behaviors that might be under consideration. In such rare cases, the problems discussed here may be substantially mitigated, though the problems caused by unreliability remain. Therefore, the present recommendation is not that researchers should never

use mediation, only that they should not draw seemingly definitive conclusions from them absent a strong prior theory. In the absence of a strong prior theory, as we have seen, too many problems plague the interpretation of beta weights and path coefficients. Nor does careful use of causally neutral language in the results section importantly mitigate the problems. The fact that arrows are used in causal diagrams renders clear that the reader is supposed to assume causation, however careful the language in the results section. And even under the pretense that the arrows in causal diagrams do not give the show away, surely the recommendations for application in discussion sections do. To ask a pointed and rhetorical question, why make recommendations for application if causation is not assumed?

Even when there is a strong prior theory, interpretation of causal diagrams can be problematical; but that is a topic best addressed in its own paper.

## References

- Beebe, H., Hitchcock, C., & Menzies, P. (Eds.) (2009). *The Oxford handbook of causation*. Oxford, U. K.: Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199279739.001.0001>
- Borsboom, D. (2005). *Measuring the mind: Conceptual issues in contemporary psychometrics*. Cambridge, England: Cambridge University Press. <https://doi.org/10.1017/CBO9780511490026>
- Borsboom, D., & Mellenbergh, G. J. (2002). True scores, latent variables and constructs: A comment on Schmidt and Hunter. *Intelligence*, 30(6), 505-514. [https://doi.org/10.1016/S0160-2896\(02\)00082-X](https://doi.org/10.1016/S0160-2896(02)00082-X)
- Borsboom, D., & Mellenbergh, G. J. (2003). The theoretical status of latent variables. *Psychological Review*, 110(2), 203-219. <https://doi.org/10.1037/0033-295X.110.2.203>
- Cartwright, N. (1983). *How the Laws of Physics Lie*. New York: Oxford University Press. <https://doi.org/10.1093/0198247044.001.0001>
- Cartwright, N. (1999). *The Dappled World: A Study of the Boundaries of Science*. Cambridge, U. K.: Cambridge University Press. <https://doi.org/10.1017/CBO9781139167093>
- Carver, R. P. (1993). The case against statistical significance testing, revisited. *Journal of Experimental Education*, 61(4), 287-292. <https://doi.org/10.1080/00220973.1993.10806591>
- Epstein, S. (1979). The stability of behavior: I. On predicting most of the people much of the time. *Journal of Personality and Social Psychology*, 37(7), 1097-1126. <https://doi.org/10.1037/0022-3514.37.7.1097>
- Epstein, S. (1980). The Stability of Behavior: II. Implications for Psychological Research. *American Psychologist*, 35(9), 790-806. <https://doi.org/10.1037/0003-066X.35.9.790>
- Fiedler, K., Schott, M., & Meiser, T. (2011). What mediation analysis can (not) do. *Journal of Experimental Social Psychology*, 47(6), 1231-1236. <https://doi.org/10.1016/j.jesp.2011.05.007>
- Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention and behavior: An introduction to theory and research*. Reading, MA: Addison-Wesley.
- Fishbein, M., & Ajzen, I. (2010). *Predicting and changing behavior: The reasoned action approach*. New York, NY: Psychology Press. <https://doi.org/10.4324/9780203838020>
- Grice, J. W. (2011). *Observation Oriented Modeling: Analysis of Cause in the Behavioral Sciences*. New York, NY: Academic Press.
- Grice, J. W., Cohn, A., Ramsey, R. R., & Chaney, J. M. (2015). On muddled reasoning and mediation modeling. *Basic and Applied Social Psychology*, 37(4), 214-225. <https://doi.org/10.1080/01973533.2015.1049350>
- Grice, J. W., Medellin, E., Jones, I., Horvath, S., McDaniel, H., O'lansen, C., & Baker, M. (2021). Persons as effect sizes. *Advances in Methods and Practices in Psychological Science*, 3(4), 443-455. <https://doi.org/10.1177/2515245920922982>
- Gulliksen, H. (1987). *Theory of mental tests*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Hulin, C. L., Drasgow, F., & Parsons, C. K. (1983). *Item response theory: Application to psychological measurement*. United States: Dow Jones-Irwin.
- Kline, R. B. (2015). The mediation myth. *Basic and Applied Social Psychology*, 37(4), 202-213. <https://doi.org/10.1080/01973533.2015.1049349>
- Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
- Mackie, J. L. (1974/1980). *The Cement of the Universe: A Study of Causation*. Oxford, U. K.: Oxford University Press.
- MacKinnon, D. P., Krull, J. L., & Lockwood, C. M. (2000). Equivalence of the mediation, confounding and suppression

- effect. *Prevention Science*, 1(4), 173-181. <https://doi.org/10.1023/A:1026595011371>
- Michelson, A. A., & Morley, E. W. (1887). On the relative motion of earth and luminiferous ether. *American Journal of Science, Third Series*, 34, 233-245. <https://doi.org/10.2475/ajs.s3-34.203.333>
- Mischel, W. (1968). *Personality and assessment*. New York: Wiley.
- Molenaar, P. C. M. (2004). A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. *Measurement*, 2(4), 201-218. [https://doi.org/10.1207/s15366359mea0204\\_1](https://doi.org/10.1207/s15366359mea0204_1)
- Molenaar, P. C. M. (2015). On the relation between person-oriented and subject-specific approaches. *Journal of Person-Oriented Research*, 1(1-2), 34-41. <https://doi.org/10.17505/jpor.2015.04>
- Pedhazur, E. J. (1997). *Multiple regression in behavioral research: Explanation and prediction* (3<sup>rd</sup> ed.). United States: Wadsworth.
- Raykov, T., & Marcoulides, G. A. (2011). *Introduction to psychometric theory*. London, UK: Routledge. <https://doi.org/10.4324/9780203841624>
- Saylors, R., & Trafimow, D. (in press). Why the increasing use of complex causal models is a problem: On the danger sophisticated theoretical narratives pose to truth. *Organizational Research Methods*.
- Schmidt, F. L., & Hunter, J. E. (1996). Measurement error in psychological research: Lessons from 26 research scenarios. *Psychological Methods*, 1(2), 199-223. <https://doi.org/10.1037/1082-989X.1.2.199>
- Schmidt, F. L., & Hunter, J. E. (1999). Theory testing and measurement error. *Intelligence*, 27(3), 183-198. [https://doi.org/10.1016/S0160-2896\(99\)00024-0](https://doi.org/10.1016/S0160-2896(99)00024-0)
- Schmidt, F. L., Le, H., & Oh, I-S. (2013). Are true scores and construct scores the same? A critical examination of their substitutability and the implications for research results. *International Journal of Selection and Assessment*, 21(4), 339-354. <https://doi.org/10.1111/ijsa.12044>
- Spearman, C. (1904). The proof and measurement of association between two things. *American Journal of Psychology*, 15(1), 72-101. <https://doi.org/10.2307/1412159>
- Spirtes, P., Glymour, C., & Scheines, R. (2000). *Causation, prediction, and search*. Cambridge, MA: MIT Press. <https://doi.org/10.7551/mitpress/1754.001.0001>
- Tate, C. U. (2015). On the overuse and misuse of mediation analysis: It may be a matter of timing. *Basic and Applied Social Psychology*, 37(4), 235-246. <https://doi.org/10.1080/01973533.2015.1062380>
- Thoemmes, F. (2015). Reversing arrows in mediation models does not distinguish plausible models. *Basic and Applied Social Psychology*, 37(4), 226-234. <https://doi.org/10.1080/01973533.2015.1049351>
- Trafimow, D. (2015). What if planetary scientists used mediation analysis to infer causation? *Basic and Applied Social Psychology*, 37(4), 197-201. <https://doi.org/10.1080/01973533.2015.1064290>
- Trafimow, D. (2017a). The probability of simple versus complex causal models in causal analyses. *Behavior Research Methods*, 49(2), 739-746. <https://doi.org/10.3758/s13428-016-0731-3>
- Trafimow, D. (2017b). Some implications of distinguishing between unexplained variance that is systematic or random. *Educational and Psychological Measurement*, 78(3), 482-503. <https://doi.org/10.1177/0013164417691573>
- Trafimow, D. (2019). A frequentist alternative to significance testing, p-values, and confidence intervals. *Econometrics*, 7(2), 1-14. <https://doi.org/10.3390/econometrics7020026>
- Trafimow, D., & Myüz, H. A. (2019). The sampling precision of research in five major areas of psychology. *Behavior Research Methods*, 51, 2039-2058. <https://doi.org/10.3758/s13428-018-1173-x>
- Trafimow, D., Hyman, M. R., & Kostyk, A. (2020). The (im)precision of scholarly consumer behavior research. *Journal of Business Research*, 114, 93-101. <https://doi.org/10.1016/j.jbusres.2020.04.008>
- Trafimow, D., Sheeran, P., Conner, M., & Finlay, K. A. (2002). Evidence that perceived behavioral control is a multidimensional construct: Perceived control and perceived difficulty. *British Journal of Social Psychology*, 41(1), 101-121. <https://doi.org/10.1348/014466602165081>
- Wang, C., Wang, T., Trafimow, D., Li, H., Hu, L., & Rodriguez, A. (2021). Extending the a priori procedure (APP) to address correlation coefficients. In Ngoc, Thach N., Kreinovich, V., Trung, N.D. (eds). *Data Science for Financial Econometrics. Studies in Computational Intelligence*, 898. Springer, Cham. [https://doi.org/10.1007/978-3-030-48853-6\\_10](https://doi.org/10.1007/978-3-030-48853-6_10)
- Woodside, A. G. (2013a). Proposing a new logic for data analysis in marketing and consumer behavior: Case study

- research of large-N survey data for estimating algorithms that accurately profile X (extremely high-use) consumers. *Journal of Global Scholars of Marketing Science*, 22(4), 277-289. <https://doi.org/10.1080/21639159.2012.717369>
- Woodside, A. G. (2013b). Moving beyond multiple regression analysis to algorithms: Calling for a paradigm shift from symmetric to asymmetric thinking in data analysis and crafting theory. *Journal of Business Research*, 66, 463-472. <https://doi.org/10.1016/j.jbusres.2012.12.021>
- Woodside, A. G. (2015). Constructing business-to-business marketing models that overcome the limitations in variable-based and case-based research paradigms. *Journal of Business-to-Business Marketing*, 22(1-2), 95-110. <https://doi.org/10.1080/1051712X.2015.1021589>
- Zimmerman, D. W. (1975). Probability spaces, Hilbert spaces, and the axioms of test theory. *Psychometrika*, 40, 395-412. <https://doi.org/10.1007/BF02291765>

Appendix

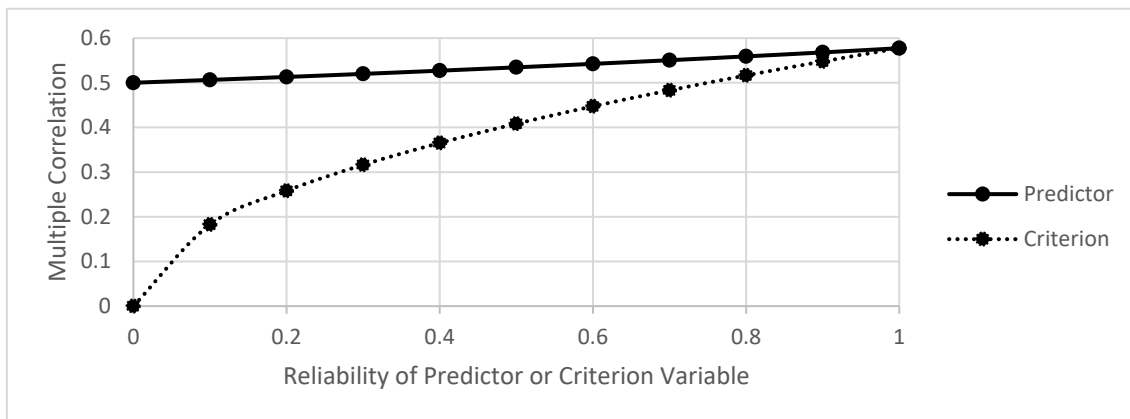


Figure 1. The multiple correlation is expressed along the vertical axis as a function of the reliability of the predictor or criterion variable along the horizontal axis (keeping other reliabilities at 1.00), where the solid curve depicts the effect of varying the reliability of a predictor variable on the multiple correlation whereas the dotted curve depicts the effect of varying the reliability of the criterion variable on the multiple correlation

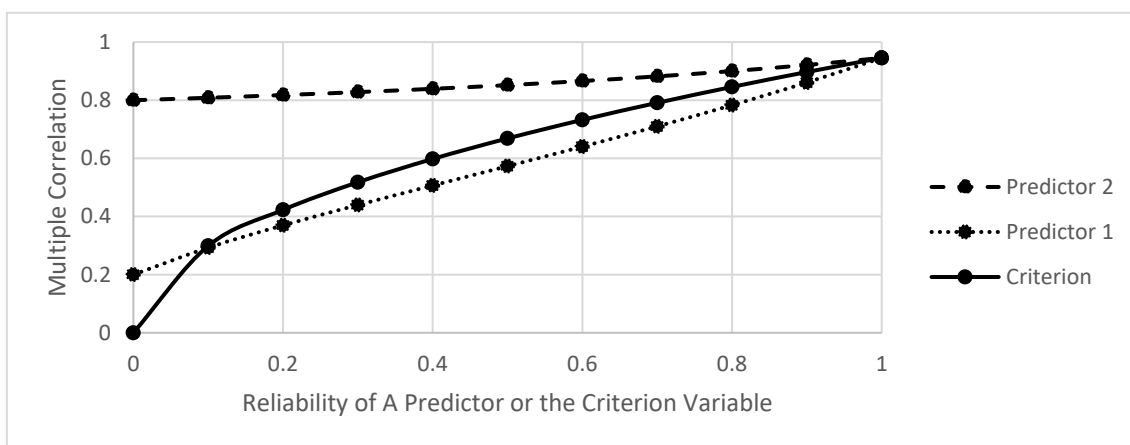


Figure 2. The multiple correlation is expressed along the vertical axis as a function of the reliability of a predictor or criterion variable along the horizontal axis (keeping the other reliabilities at 1.00), where the dashed curve depicts the effect of varying the reliability of the second predictor variable on the multiple correlation, the solid curve depicts the effect of varying the reliability of the criterion variable on the multiple correlation, and the dotted curve depicts the effect of varying the reliability of the first predictor variable on the multiple correlation

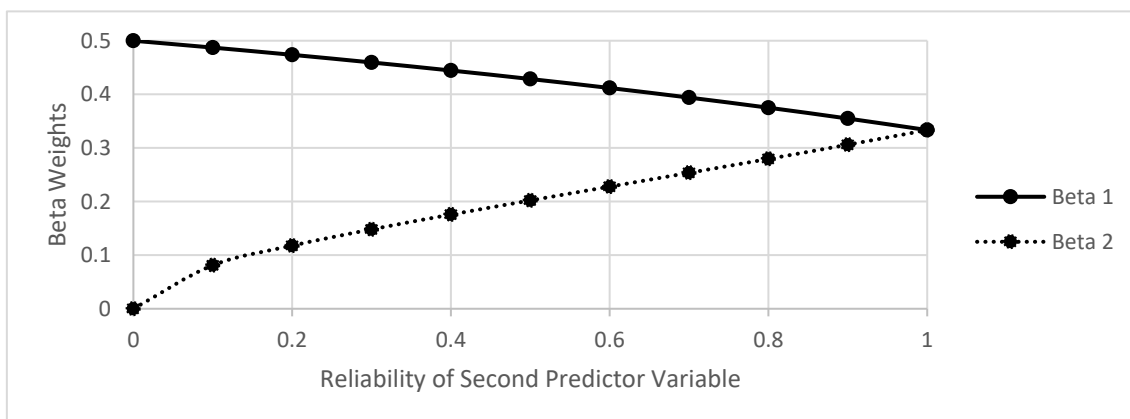


Figure 3. The beta weights are expressed along the vertical axis as a function of the reliability of the second predictor along the horizontal axis (keeping the other reliabilities at 1.00), with the solid curve representing the beta weight of the first predictor variable (Beta 1) and the dotted curve representing the beta weight of the second predictor variable (Beta 2)

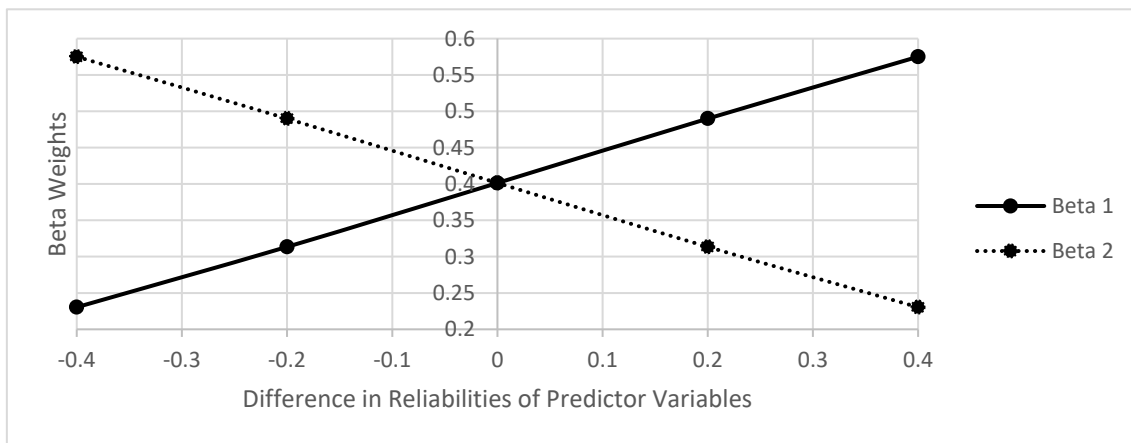


Figure 4. The beta weights are expressed along the vertical axis as a function of the difference in reliabilities of the predictor variables along the horizontal axis, with the solid curve representing the beta weight of the first predictor variable (Beta 1) and the dotted curve representing the beta weight of the second predictor variable (Beta 2)

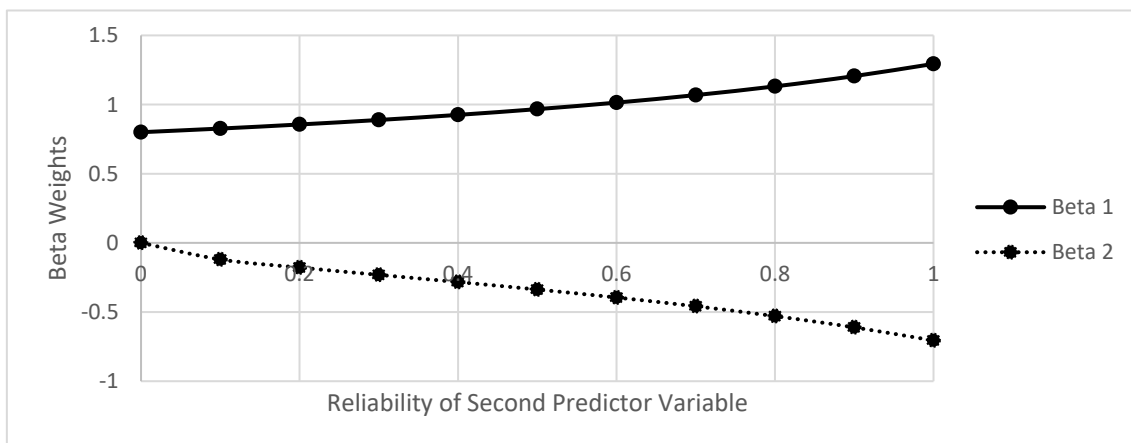


Figure 5. The beta weights are expressed along the vertical axis as a function of the reliability of the second predictor variable along the horizontal axis (keeping the other reliabilities at 1.00), with the solid curve representing the beta weight of the first predictor variable (Beta 1) and the dotted curve representing the beta weight of the second predictor variable (Beta 2)

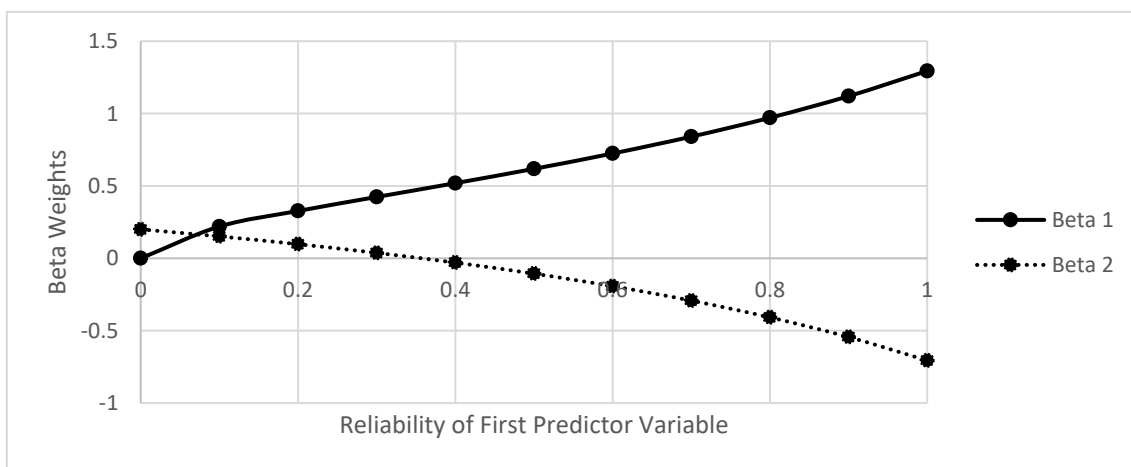


Figure 6. The beta weights are expressed along the vertical axis as a function of the reliability of the first predictor variable along the horizontal axis (keeping the other reliabilities at 1.00), with the solid curve representing the beta weight of the first predictor variable (Beta 1) and the dotted curve representing the beta weight of the second predictor variable (Beta 2)

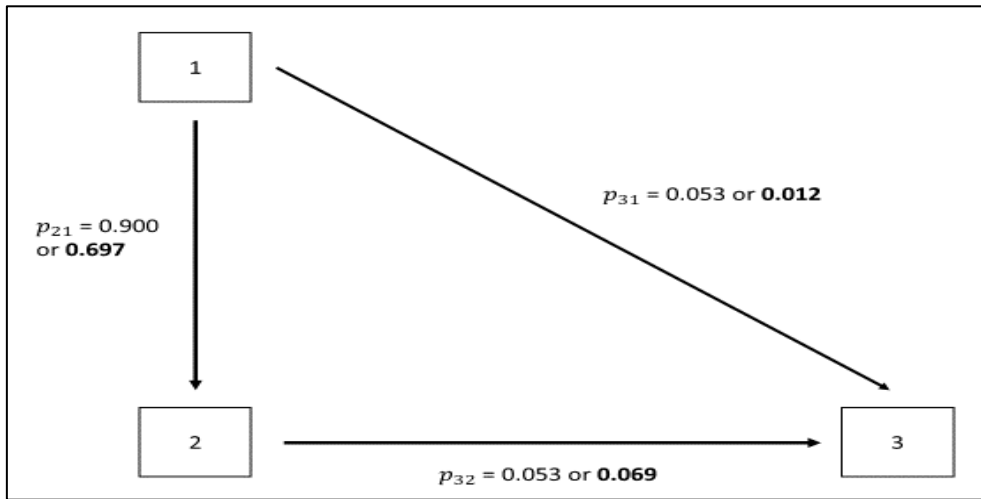


Figure 7. Path diagram showing how the direct pathway from Variable 1 to Variable 3 almost disappears when their reliabilities decrease from 1.0 to 0.6

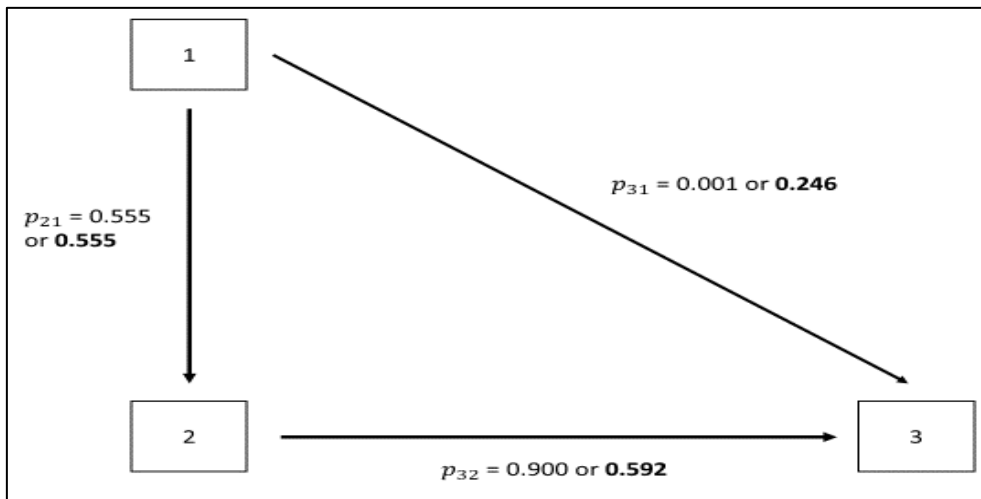


Figure 8. Path diagram showing how the direct pathway from Variable 1 to Variable 3 appears out of thin air when the reliability of Variable 2 decreases from 1.0 to 0.6

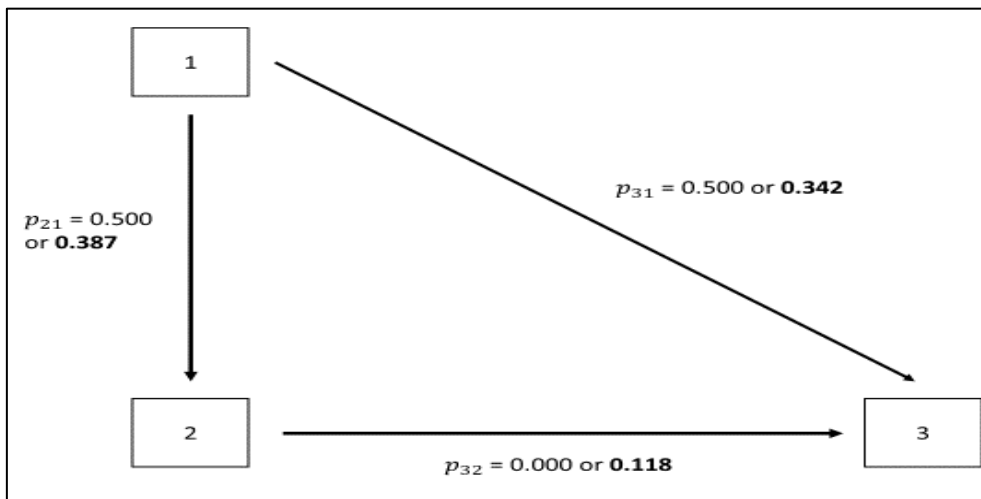


Figure 9. Path diagram showing how the pathway from Variable 2 to Variable 3 appears out of thin air when the reliability of Variable 1 decreases from 1.0 to 0.6

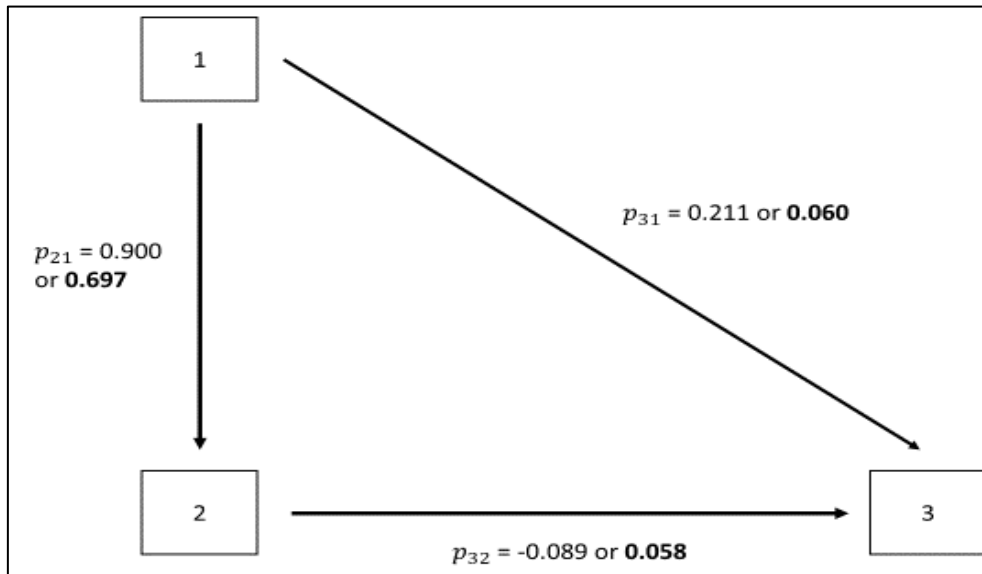


Figure 10. Path diagram showing how the pathway from Variable 2 to Variable 3 changes sign when the reliability of Variable 1 decreases from unity to 0.6

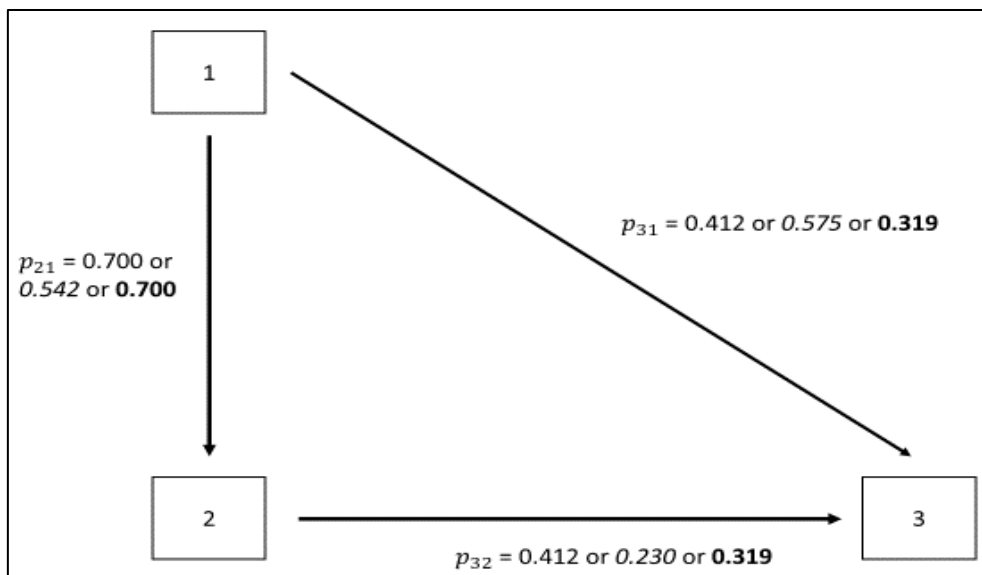


Figure 11. Path diagram showing how the unreliability can cause a switching of mediators, with the italicized values indicating when the reliability of Variable 2 is set at 0.60 and the boldfaced values indicating when the reliability of Variable 3 is set at 0.60



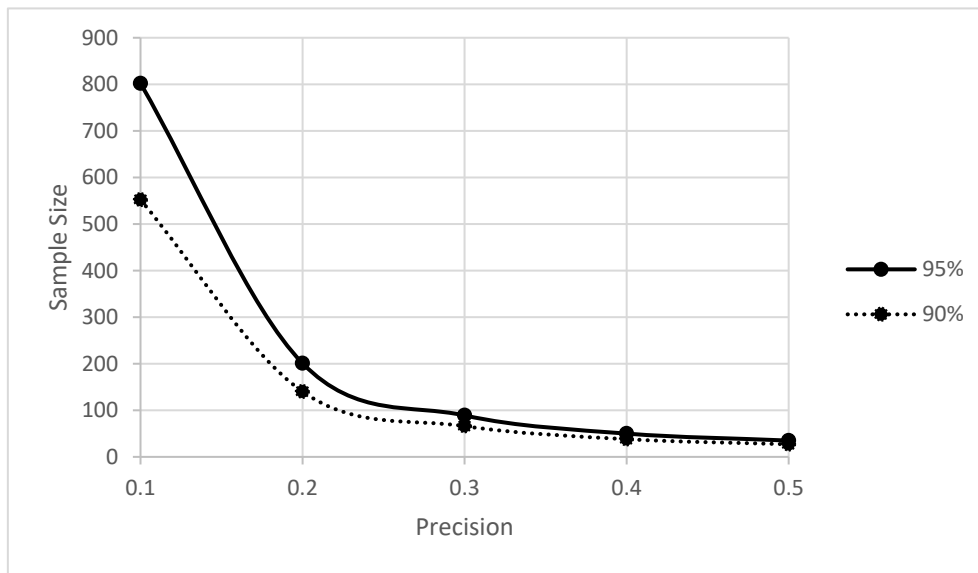


Figure 12. Sample size varies along the vertical axis as a function of precision along the horizontal axis, at 95% confidence (solid curve) or 90% confidence (dotted curve)

**Copyrights**

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the [Creative Commons Attribution license](#) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.