T. A., & Janowsky, J. S. (2010). Assessment of body image in younger and older men. *Journal of General Psychology*, *137*, 225–238.
:ts, T. A. (2004). Female trouble: The menstrual self-evaluation scale and women's :-objectification. *Psychology of Women Quarterly*, *28*, 22–26.
:ts, T. A., & Waters, P. L. (2004). Self-objectification and that "not so fresh feeling": ninist therapeutic interventions for healthy female embodiment. *Women & Therapy*, 5–21.
k, A., Engeln-Maddox, R., & Miller, S. A. (2010). Here's looking at you: Self-ectification, body image disturbance, and sorority rush. *Sex Roles*, *63*(1–2), 6–17.
:al, M., Steyn, N., Mashego, T., & Nel, J. (2001). Evaluation of body shape, eating orders and weight management related parameters in black female students of rural l urban origins. *South African Journal of Psychology*, *31*, 45–53.
er, D., Lupfer-Johnson, G., & Rivkin, I. (2008). Yoga as an intervention for self-ectification. *Psi Chi Journal of Undergraduate Research*, *13*, 168–172.
ak, L., & Murnen, S. K. (2008). Drive for leanness: Assessment and relationship to ıder, gender role, and objectification. *Body Image*, *5*, 251–260.
on, J. M. (1998). An empirical assessment of data collection using the Internet. *sonal Psychology*, *51*, 709–725.
E., Marti, C., Shaw, H., & Jaconis, M. (2009). An 8-year longitudinal study of the ural history of threshold, subthreshold, and partial eating disorders from a community nple of adolescents. *Journal of Abnormal Psychology*, *118*(3), 587–597.
, J. K., Hyers, L. L., Cohen, L. L., & Ferguson, M. J. (2001). Everyday sexism: idence for its incidence, nature, and psychological impact from three daily diary dies. *Journal of Social Issues*, *57*, 31–53.
ıpson, J., & Heinberg, L. J. (1999). The media's influence on body image disturbance l eating disorders: We've reviled them, now can we rehabilitate them? *Journal of cial Issues*, *55*(2), 339–353.
ıpson, J. K., & Stice, E. (2001). Thin-ideal internalization: Mounting evidence for a v risk factor for body-image disturbance and eating pathology. *Current Directions in ychological Science*, *10*, 181–183.
ıpson, J. K., van den Berg, P., Roehrig, M., Guarda, A. S., & Heinberg L. J. (2004). e Sociocultural Attitudes Toward Appearance Scale-3 (SATAQ-3): Development and idation. *International Journal of Eating Disorders*, *35*, 293–304.
mann, M., & Lynch, J. E. (2001). Body image across the life span in adult women: e role of self-objectification. *Developmental Psychology*, *37*, 243–253.

Original manuscript received December 18, 2010
Final version accepted February 7, 2011

The Journal of General Psychology, 2011, *138*(3), 201–214

An Investigation of the Accuracy of Standardized Path Coefficients

DAVID TRAFIMOW
STEPHEN RICE
JUSTIN A. MACDONALD
New Mexico State University

ABSTRACT. Given the popular and ever-increasing use of path analytic research paradigms in the social sciences, it is desirable to conduct an investigation into the accuracy of the standardized path coefficients that are often the end-product of these paradigms. In pursuit of this goal, population parameters were preset concerning the correlations between all of the variables and their reliability coefficients. Based on these parameters, thousands of experiments were generated with varying numbers of cases (n). For each experiment, at each level of n, standard path analyses were conducted, and standardized path coefficients were obtained. These standardized path coefficients were then compared against the population path coefficients on which the simulations were based to determine their accuracy. The findings indicate mixed evidence for the accuracy of path analysis research paradigms.

Keywords: path analyses, reliability coefficients, simulation, standardized path coefficients

PATH ANALYTIC RESEARCH PARADIGMS have become increasingly popular in the social sciences, at least in part because they serve a variety of potential functions. For example, several authorities have indicated that, properly used, path analytic paradigms enable researchers to draw causal conclusions form correlational data (Angrist, Imbens, & Rubin, 1996; Hope, 1984; Kenny, 1979; Mulaik, 1987). In addition, these paradigms allow researchers to handle a large number of correlated variables in one set of related equations (James & Brett, 1984; MacKinnon, 2000; MacKinnon, Krull, & Lockwood, 2000; Pedhazer, 1997). The combination of these two capabilities, in turn, confers on researchers the ability to determine complex patterns of causation, including mediation and moderation. The totality of these capabilities, without the requirement of potentially less practicable experiments, provides a strong argument for the utility of path analytic paradigms (McClendon, 1994).

Address correspondence to Stephen Rice, Department of Psychology, MSC 3452, New Mexico State University, PO Box 30001, Las Cruces, NM 88003-8001, USA; sc_rice@yahoo.com (e-mail).

Nevertheless, there remain critics who are not convinced of the promised benefits. For example, several researchers have pointed out that path analyses depend on correlations, and one cannot make a strong case for causation on that basis (Freedman 1987; Holland, 1986; Rogosa, 1987). However, path analysis aficionados have countered that the idea is not to infer causation but rather to test alternative causal models against each other. If a path analysis is more consistent with one causal model than another, then that is a reason to favor the former model over the latter one. Another criticism that has been leveled against path analytic research paradigms is that they require assumptions that are unlikely to be true, although defenders contend that the assumptions are not as bad as critics make them out to be (for discussions, see Maruyama, 1998; McClendon, 1994; Pedhazur, 1997). Critics also have pointed out that because social science measures are not perfectly reliable and because unreliability has unpredictable effects on obtained path coefficients (Blalock, 1964; Cohen & Cohen, 1983; Liu, 1988), the obtained (uncorrected) path coefficients cannot be trusted. However, this criticism can be countered by the claim that correlations can be corrected for attenuation due to unreliability. If the path analyses are based on corrected correlations, then the criticism no longer applies (for a review, see Trafimow, 2006a).

Although we are interested in the various arguments and counterarguments that pertain to path analysis research paradigms, that is not our present focus. Rather, our strategy is to assume that path analytic research paradigms are philosophically justifiable when they are based on corrected correlations and address a different issue. Specifically, how accurate are standardized path coefficients when they are based on corrected correlations? Obviously, to address this issue, knowledge of the population path coefficients is necessary. Although this knowledge is generally unobtainable in real research, it is possible to perform simulations based on population parameters that are preset. Using these preset parameters, it is possible to have the computer generate thousands of studies based on them, perform path analyses, and compare the obtained standardized path coefficients against the population parameters on which they were generated.

There is, of course, an important complication. When correlations are corrected for attenuation due to unreliability, the reliability coefficients obtained in any particular study are not exactly correct; there is a sampling distribution of reliability coefficients just as there is a sampling distribution of other statistics. Thus, although the reliability coefficients obtained in particular studies can be used to estimate population reliability coefficients, the fact that these are estimates, rather than actual population parameters, introduces an additional source of error. Thus, there is error associated with the obtained correlations, and there is error associated with the reliability coefficients that are used to correct them.

Many researchers who perform path analyses fail to adjust the correlations for attenuation due to unreliability, which is just plain wrong. The more sophisticated path modelers do make the corrections but nevertheless tend not to focus on the fact that the reliability coefficients themselves are a source of error. Thus, there is a reason to suspect that path analyses might be less accurate than researchers typically assume because the contribution of error associated with using reliability coefficients has not received much press. Furthermore, the interaction between the two classes of errors has not been explored but might further decrease the accuracy of standardized path coefficients. It seems obvious that if researchers are going to interpret standardized path coefficients, it is important to have some idea of their accuracy and, if proven to be inaccurate, confidence in such interpretations should decrease accordingly.

To test these ideas, we performed computer simulations using preset population parameters for reliability coefficients as well as for correlation coefficients.

Method

The Simulation Strategy

The first step was to preset the population parameters. Among other issues, this meant making arbitrary decisions about what the population reliability coefficients would be and what the population standardized path coefficients and correlations would be. We preset the population reliability coefficients of all of the variables at .8. Also for simplification, we used only three variables; predictor variable X, potential mediating variable M, and criterion variable Y. Thus, there were only three correlations from which all standardized path coefficients could be constructed: (a) the correlation between X and M (r_{XM}), (b) the correlation between X and Y (r_{XY}), and (c) the correlation between M and Y (r_{MY}). The generated sample correlation coefficients for one set of simulations were based on the following population correlation coefficients: (a) ρ_{XY} was preset at 0.4, (b) ρ_{XM} was preset at 0.4, and (c) ρ_{MY} was preset at 0.16. The population coefficients used in the second set of simulations were ρ_{XY} preset at 0.5, ρ_{XM} preset at 0.5, and ρ_{MY} preset at 0.25. The population correlation coefficients used in the third set of simulations were ρ_{XY} preset at 0.6, ρ_{XM} preset at 0.6, and ρ_{MY} preset at 0.36. The population coefficients used in the fourth set of simulations were ρ_{XY} preset at 0.7, ρ_{XM} preset at 0.9, and ρ_{MY} preset at 0.63. Note that with these population correlation coefficients, the population standardized path coefficient from M to Y is zero in all four sets of simulations, and so there is no mediation in the population in either case. By using zero, we hoped to make the errors particularly easy to perceive. The issue of interest in both simulations was how close the generated path coefficients from M to Y would be to the true value of zero.

To proceed with the simulation it is necessary to generate observed scores from which correlations (r_{XM}, r_{XY}, and r_{MY}) and reliability estimates can be obtained. In accordance with classical test theory, an observed score is the sum of the true score and an error term. True scores were generated by sampling from a trivariate normal distribution with mean vector $\mu = [\mu_X\, \mu'_M\, \mu_Y]$ and one of the following covariance matrixes:.

$$\sum_1 = \begin{bmatrix} 1 & 0.4 & 0.4 \\ 0.4 & 1 & 0.16 \\ 0.4 & 0.16 & 1 \end{bmatrix} \text{ or } \sum_2 = \begin{bmatrix} 1 & 0.9 & 0.7 \\ 0.9 & 1 & 0.63 \\ 0.7 & 0.63 & 1 \end{bmatrix}$$

Error terms sampled from a normal distribution with mean zero and variance 0.25 were added to each coordinate of the true score to obtain an observed score for the simulation. An error variance of 0.25 was chosen so that the population reliability coefficients would be equal to 0.8. A total of 1,700,000 studies were generated assuming sample sizes of 50, 100, 200, 500, 1,000, 1,500, ..., 5,000 (65,000 for each sample size). In each study, r_{XY}, r_{XM}, and r_{MY} were calculated from the sample of observed scores. The test–retest method was used to obtain reliability estimates (r_{XX}', r_{MM}', and r_{YY}') for each sample: Another set of observed scores was generated from the underlying set of true scores and reliability was estimated by correlating the scores across tests.

Given that the previous step resulted in r_{XM}, r_{XM}, and r_{XM}, as well as r_{XX}', r_{MM}', and r_{YY}', the next step was to correct the three correlation coefficients using the three reliability coefficients. This was done according to Equations 1, 2, and 3 below. We used upper-case letters to indicate the best estimates of population parameters so that R_{XM} is the best estimate of ρ_{XM}, R_{MY} is the best estimate of ρ_{MY}, and R_{XY} is the best estimate of ρ_{XY}.

$$R_{XM} = \frac{r_{XM}}{\sqrt{r_{XX'} r_{MM'}}} \tag{1}$$

$$R_{MY} = \frac{r_{MY}}{\sqrt{r_{MM'} r_{YY'}}} \tag{2}$$

$$R_{XY} = \frac{r_{XY}}{\sqrt{r_{XX'} r_{YY'}}} \tag{3}$$

Given the corrected correlations, R_{XM}, R_{MY}, and R_{XY}, it was now possible to compute standardized path coefficients according to Equations 4, 5, and 6. Because these were based on corrected correlations, upper-case letters will be used, indicating that these standardized path coefficients are the best estimates that can be made from the corrected sample data. Note that it is the last equation—Equation 6—that we focus on as the population standardized path coefficient equals zero for all four sets of simulations.

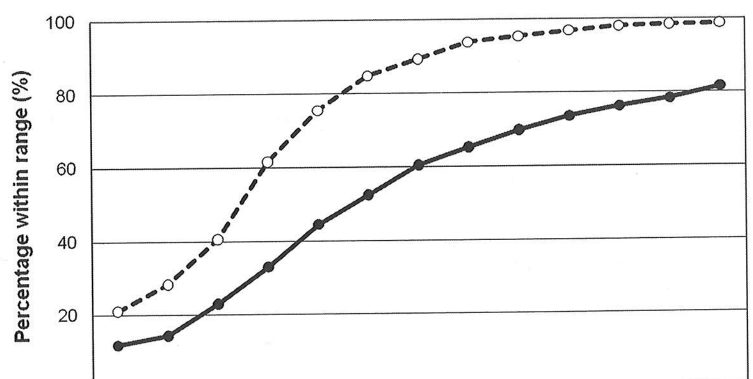
$$P_{XM} = R_{XM} \tag{4}$$

$$P_{XY} = \frac{(R_{XY} - R_{MY} R_{XM})}{(1 - R_{XM}^2)} \tag{5}$$

$$P_{MY} = \frac{(R_{MY} - R_{XY} R_{XM})}{(1 - R_{XM}^2)} \tag{6}$$

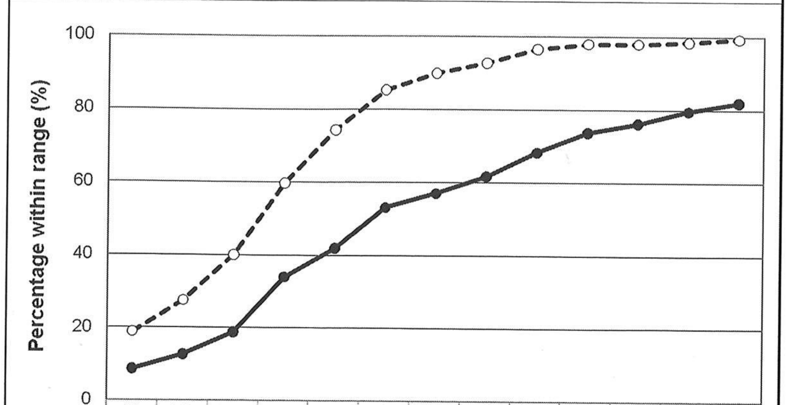
Results

Given that the preset value of the standardized path coefficient from M to Y always was zero, the mean across all of the simulations should have approximately equaled zero and a failure to obtain this result would indicate that we had done something wrong in the simulations. Thus, before looking at the findings of main interest, we checked this mean, and it was essentially zero. In addition, all of the mean standardized path coefficients were essentially equal to the preset values, thereby indicating that the simulations were performed correctly. Let us now address our main interest—how often do the standardized path coefficients in studies approximate the actual population coefficients? Figures 1–4 answer this question. Figure 1 is based on the set of smallest population correlation coefficients ($\rho_{XY} = 0.4$, $\rho_{XM} = 0.4$, and $\rho_{MY} = 0.16$), Figure 2 is based on the second set of population correlation coefficients ($\rho_{XY} = 0.5$, $\rho_{XM} = 0.5$, and $\rho_{MY} = 0.25$), Figure 3 is based on the third set of population correlation coefficients ($\rho_{XY} = 0.6$, $\rho_{XM} = 0.6$, and $\rho_{MY} = 0.36$), and Figure 4 is based on the largest set of population correlation coefficients ($\rho_{XY} = 0.7$, $\rho_{XM} = 0.9$, and $\rho_{MY} = 0.63$). In all four figures, we computed the percentage of times that the sample standardized path coefficient indicating the strength of the path between M and Y (P_{MY}) was within 0.05 ($\pm$ 0.025) or 0.1 ($\pm$ 0.05) of the population standardized path coefficient (remember that this was preset at zero). When the sample size is a relatively small but



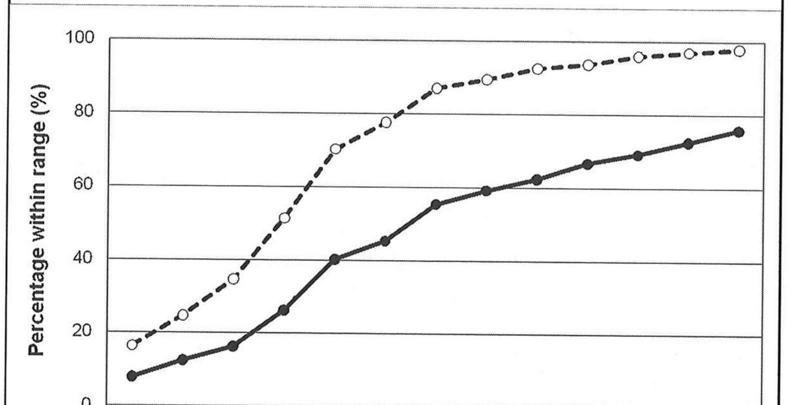
	50	100	200	500	1,000	1,500	2,000	2,500	3,000	3,500	4,000	4,500	5,000
0.1	21.1	28.3	40.6	61.6	75.5	84.9	89.4	93.9	95.4	96.9	98.1	98.5	98.8
0.05	11.8	14.4	23	33	44.5	52.5	60.5	65.3	69.8	73.6	76.3	78.4	81.6

FIGURE 1. Four correlation between X and M.



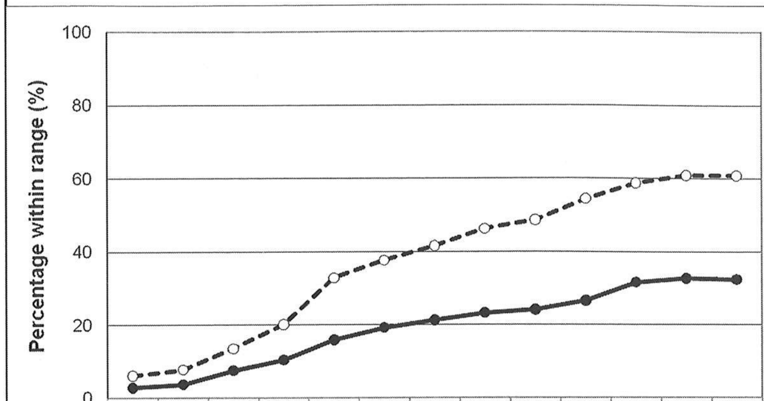
	50	100	200	500	1,000	1,500	2,000	2,500	3,000	3,500	4,000	4,500	5,000
0.1	18.9	27.5	40.1	59.8	74.3	85.2	89.8	92.7	96.5	97.9	97.9	98.4	99.3
0.05	8.7	12.7	18.8	34	42	53.2	57	61.7	68.3	73.7	76.2	79.6	82

FIGURE 2. Five correlation between *X* and *M*.



	50	100	200	500	1,000	1,500	2,000	2,500	3,000	3,500	4,000	4,500	5,000
0.1	16.4	24.7	34.6	51.4	70.3	77.7	87.1	89.6	92.6	93.8	96.1	97.1	98
0.05	7.9	12.5	16.3	26.2	40.2	45.3	55.4	59.2	62.4	66.7	69.2	72.5	75.9

FIGURE 3. Six correlation between *X* and *M*.



	50	100	200	500	1,000	1,500	2,000	2,500	3,000	3,500	4,000	4,500	5,000
0.1	6.2	7.8	13.6	20.2	32.9	37.7	41.6	46.3	48.7	54.5	58.8	60.9	60.9
0.05	3	3.8	7.6	10.5	16	19.3	21.3	23.3	24.2	26.7	31.7	32.8	32.5

FIGURE 4. Nine correlatioun between *X* and *M*.

realistic 50, Figure 1 shows that only 11.8% of the obtained standardized path coefficients are within 0.05 of the true value and, even with an extremely liberal criterion of 0.1, this percentage increases to only 21.1%. Increasing the sample size to 2,500 increases accuracy such that 65.3% of the standardized path coefficients are within 0.05 of the true value and 93.9% are within 0.1 of the true value. Finally, when the sample size increases to 5,000, a more impressive 81.6% of the obtained standardized path coefficients are within 0.05 of the true value, and an even more impressive 98.8% are within 0.1 of the true value. Thus, Figure 1 is a generally pessimistic illustration of the accuracy of path analyses unless extremely large sample sizes are used.

Figures 2–4, which are based on increasingly larger population correlations, suggest even more pessimism. At worst, consider Figure 4. When the sample size is 50, only 3.0% of the standardized path coefficients are within 0.05 of the true value and this only increases to 6.2% when a 0.1 criterion is used. Increasing the sample size to 2500 improves accuracy such that 23.3% of the standardized path coefficients are within 0.05 of the true value and 46.3% meet the 0.1 criterion. Even when the sample size is increased further to 5,000, only 32.5% meet the 0.05 criterion and 60.9% meet the 0.1 criterion.

An alternative way to explore the accuracy issue is to compute the standard deviations of the standardized path coefficients. Table 1 presents these for all four

TABLE 1. Standard Deviations of Standardized Path Coefficients Between M and Y as a Function of the Sizes of the Population Correlation Coefficients and the Sample Sizes

	Sample sizes												
Correlation	50	100	200	500	1,000	1,500	2,000	2,500	3,000	3,500	4,000	4,500	5,000
0.9	7.31	3.94	0.55	0.20	0.13	0.11	0.09	0.08	0.08	0.07	0.06	0.06	0.06
0.6	0.25	0.16	0.12	0.07	0.05	0.04	0.03	0.03	0.03	0.03	0.02	0.02	0.02
0.5	0.21	0.14	0.10	0.06	0.04	0.03	0.03	0.03	0.02	0.02	0.02	0.02	0.02
0.4	0.21	0.14	0.10	0.06	0.04	0.03	0.03	0.03	0.02	0.02	0.02	0.02	0.02

sets of simulations. Consistent with the foregoing simulations, standard deviations decrease as the population correlations decrease and as the sample sizes increase.

Discussion

The results can be summarized easily. Of greatest importance, although the population correlations were set up in such a way so as to make the standardized path coefficient from M to Y equal to zero (no mediation), it was a rare occasion when the results from a particular study were consistent with this. Rather, the standardized path coefficients obtained in particular studies often failed to be within 0.05, or even 0.1, of the correct value even when fairly large sample sizes were used. Thus, mediation was likely to be demonstrated even when there was none in the population. Not surprisingly, accuracy increased as the sample size increased and accuracy was also influenced by whether the population correlations were larger or smaller. Note that these findings were obtained with correlation coefficients that were corrected for attenuation due to the unreliability of the measures. Had we not corrected for attenuation, the accuracy of the path analyses would have been found to be substantially decreased relative to what Figures 1–4 demonstrate.

One potential interpretation of our simulations is that they provide no reason for worry. Because accuracy increases as sample size increases, it is merely necessary to use a large sample size and then the universe is friendly. We see three problems with this interpretation. The first problem is a practical one. Specifically, our minimum sample size was 50, which is fairly common in path analytic studies, and Figures 1 and 2 showed that many more participants than that are needed to obtain a reasonable degree of accuracy. Interestingly, we have been informed on more than one occasion by researchers who use the paradigm that a major inducement for doing so is precisely because one does not need many participants.[1] In contrast, Figures 1–4 demonstrate that very large sample sizes are critical, or the obtained path coefficients are likely to be inaccurate. It is possible that many of the mediation findings that have been obtained in the psychological literature are spurious.

This pessimistic conclusion actually may be overly optimistic when a second issue is considered. Specifically, our simulations were performed with only three variables. What if many more variables are considered, as is generally the case when researchers use path analytic paradigms? In this case, there would be yet further sources of error because there are more correlations to obtain, and because there are more reliability coefficients to obtain. As these additional sources of error are included, the number of participants necessary for reasonable accuracy is likely to increase.

Then, too, a third factor to consider is that we made the very generous assumption that the only errors are random errors and that there are no nonrandom errors. This assumption is practically certain to be false in real research, and

Trafimow (2006a) has documented the deleterious effects of nonrandom invalidity on obtained standardized path coefficients. The combination of all three of these factors suggests that path analytic paradigms should be used with extreme caution.

It is possible to imagine potential objections. For example, one might argue that statistical significance tests take care of the problem of inferring mediation when there is none, and so the inaccuracy demonstrated in Figures 1 and 2 is not so important. But this potential objection can be refuted easily. For one thing, it is well known that obtained p values fail to provide the probability of wrongly rejecting the null hypothesis (Bakan, 1966; Carver, 1978, 1993; Cohen, 1994; Meehl, 1978; Nickerson, 2000; Rozeboom, 1960, 1997; Schmidt, 1996; Schmidt & Hunter, 1997, 2002; Trafimow, 2003; for a recent review, see Trafimow, 2006b). This is because, contrary to popular belief, p is *not* the probability that the null hypothesis is true given the finding but rather is the probability of obtaining a finding in a particular range given that the null hypothesis is true (Nickerson, 2000). Even when p is an extremely low number (e.g., less than 0.05), the probability of the null hypothesis given the finding can be anywhere between 0 and 1 (see Trafimow, 2003, Figure 1; for a review, see Trafimow, 2006b).[2] In addition, Trafimow and Rice (2009) demonstrated that the correlation between p values and the probabilities of the hypotheses to which they correspond is low and becomes even lower when p values are used to make dichotomous decisions about hypotheses.

It is interesting to contrast Figures 1–4, and the contrast between Figure 1 and Figure 4 is particularly easy to perceive. Although Figure 1 indicates that accuracy is low unless extremely large sample sizes are used, Figure 4 indicates that, even when the sample size is 5,000, accuracy nevertheless remains poor. What is the difference between Figures 1 and 4 telling us? To answer this question, consider again that Figure 1 was based on generally small population correlation coefficients ($\rho_{XY} = 0.40$, $\rho_{XM} = 0.40$, and $\rho_{MY} = 0.16$), whereas Figure 4 was based on larger ones ($\rho_{XY} = 0.70$, $\rho_{XM} = 0.90$, and $\rho_{MY} = 0.63$). Clearly, then, it seems that larger correlation coefficients lead to more inaccuracy than do small ones (note that Figures 2 and 3 are intermediate), although it is important to remember that the means of the sample standardized path coefficients are appropriately very close to the population value of zero regardless of whether small or large correlations population correlation coefficients were used. A possible explanation is that higher correlations mean more multicollinearity, thereby reducing the stability of the regression coefficients.

It could be argued that this is a good thing. Many correlations in the social science literature tend to be small or moderate (e.g., between .2 and .6), and so it is better if small correlations give more accurate M to Y path coefficients than if larger ones do. But this argument fails to take into account the difference between correlation coefficients that have been corrected for attenuation due to the unreliability of the measures versus correlations coefficients that have not been so corrected. To understand why this is important, consider Equation 7 below, which was obtained simply by rearranging the terms from Equation 6.

$$r_{XY} = R_{XY}\sqrt{r_{XX'}r_{YY'}} \tag{7}$$

Suppose that we instantiate some numbers into Equation 7 keeping the "true" or corrected correlation constant at 0.7. First, suppose that both measures are reliable at the 0.9 level, in which case the expected obtained correlation would be 0.63, which is not dramatically far away from the correct value of 0.7. But the decrease becomes increasingly more dramatic as the measures are set at less reliable levels such as 0.8 (expected $r_{XY} = 0.56$), 0.7 (expected $r_{XY} = 0.49$), 0.6 (expected $r_{XY} = 0.42$), and so on. Consequently, Equation 7 provides reason to suspect that the obtained correlations in the literature are substantially less than what they would be if corrected using Equation 6. In turn, this implies that the true or corrected correlations between variables common in the literature really are large, in which case the pessimistic implications of Figure 4 retain their full force.

But why are obtained standardized path coefficients so inaccurate, especially given that we only used three variables? To answer this question, note that there were actually six, and not three, sources of error. That is, there are sampling distributions associated not only with the population correlations of interest (ρ_{XM}, ρ_{MY}, and ρ_{XY}) but also with the reliability coefficients that are used to correct them (ρ_{XX}', ρ_{MM}', and ρ_{YY}'). With all of the sources of error, even if each source contributes only a small amount of error, the multiplication of errors quickly decreases the obtained accuracy. To make salient the extent of such multiplication, consider that instead of using four equations to obtain an estimate of P_{MY}; three to correct the relevant correlation coefficients and one where they can be instantiated to obtain the estimate of P_{MY}; it is possible to combine these operations in a single equation, as is shown by Equation 8. Equation 8 includes all six sources of error to illustrate how they combine to contribute to inaccuracy. Of course, as we indicated earlier, as more variables are added, there are more correlations of interest and more reliability coefficients, thereby further decreasing accuracy. For example, if there are four variables (X, M, Q, and Y), then there are six correlation coefficients of concern (ρ_{XM}, ρ_{XQ}, ρ_{XY}, ρ_{MQ}, ρ_{MY}, and ρ_{QY}) and four reliability coefficients of concern (ρ_{XX}', ρ_{MM}', ρ_{QQ}', and ρ_{YY}'), making for $6 + 4 = 10$ sources of error rather than the six in the present simulations.

$$P_{MY} = \frac{\left(\frac{r_{MY}}{\sqrt{r_{MM'}r_{YY'}}} - \frac{r_{XY}}{\sqrt{r_{XX'}r_{YY'}}} - \frac{r_{XM}}{\sqrt{r_{XX'}r_{MM'}}}\right)}{\left(1 - \frac{r_{XM}^2}{r_{XX'}r_{MM'}}\right)} \tag{8}$$

The foregoing simulations indicate that researchers should be more cautious than they currently are about interpreting standardized path coefficients. The unpleasant fact of the matter is that standardized path coefficients based on typical sample sizes are not very accurate. In addition, the accuracy decreases rather than increases when the base level correlations increase. We think it would be a useful exercise for researchers to obtain a matrix not just of observed correlations but of corrected correlations too, to get an estimate of how large the base correlations actually are. This can be accomplished via Equations 1–3. If the corrected correlations are large, and we suspect they would be in many cases, the researcher would have even more reason to be cautious about interpreting the standardized path coefficients. Researchers should present corrected correlation matrices as standard reporting practice so readers can make their own judgments about how much confidence to place in the standardized path coefficients.

In conclusion, even assuming the truth of all of the assumptions that path analysis aficionados need to make, the present simulations demonstrate that there are major accuracy problems unless sample sizes are employed that vastly exceed those typically used in psychology publications. The simulations call into question the majority of this literature and suggest that future researchers should either use much larger sample sizes or be more cautious in making inferences about mediation.

NOTES

1. Citations are omitted to protect the guilty.
2. The posterior probability of the null hypothesis (probability of the null given the finding) clearly depends, in part, on its prior probability. In turn, the prior probability depends on whether the null hypothesis specifies a point or a range (for a review, see Trafimow 2006b). If the null hypothesis specifies a point, then its prior probability approaches zero thereby rendering the whole issue of its rejection on the basis of obtained data to be moot. However, if the null hypothesis specifies a range, so that it has a respectable (nonzero) prior probability, then Figure 1 from Trafimow (2003) demonstrates that it may have a large posterior probability too, even when p is low (e.g., less than .05). More generally, Trafimow (2003; Figure 1) demonstrates that when a null hypothesis is set up that has a reasonable prior probability of being true then the null hypothesis significance testing procedure cannot validly be used to reject it. And when the prior probability of the null hypothesis is extremely low, then it can be more validly rejected but rejecting a hypothesis that is known, prior to acquiring the data, to probably be false, provides for little gain in knowledge (Trafimow, 2003; Figure 2). Thus, the null hypothesis significance testing procedure provides a poor basis for drawing conclusions about the probabilities of hypotheses in general, or more specifically for hypotheses about mediation.

AUTHOR NOTES

David Trafimow is a professor of psychology at New Mexico State University, a fellow of the Association for Psychological Science, and an executive editor of the *Journal of General Psychology*. He received his PhD in psychology from the University of Illinois at Urbana-Champaign in 1993. His current research interests include attribution, attitudes, crosscultural research, methodology, and potential performance theory. **Stephen Rice** is an associate professor of psychology at New Mexico State University and a consulting editor of the *Journal of General Psychology*. He received his PhD in psychology from the University of Illinois at Urbana-Champaign in 2006. His current research interests include automation, trust, performance, and aviation psychology. **Justin A. MacDonald** is an assistant professor of psychology at New Mexico State University. He received his PhD in psychology at Purdue University in 2003. His current research interests include auditory perception and statistical modeling of decision-making processes.

REFERENCES

Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association, 91*, 444–472.
Bakan, D. (1966). The test of significance in psychological research. *Psychological Bulletin, 66*, 423–437.
Blalock, H. M. (1964). *Causal inferences in nonexperimental research.* Chapel Hill, NC: University of North Carolina Press.
Campbell, D. T., & Stanley, J. C. (1966). *Experimental and quasi-experimental designs for research.* Chicago, IL: Rand McNally.
Carver, R. P. (1978). The case against statistical significance testing. *Harvard Educational Review, 48*, 378–399.
Carver, R. P. (1993). The case against statistical significance testing, revisited. *Journal of Experimental Education, 61*, 287–292.
Cohen, J. (1994). The earth is round ($p < .05$). *American Psychologist, 49*, 997–1003.
Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences* (2nd Ed.). Hillsdale, NJ: Erlbaum.
Freedman, D. A. (1987). As others see us: A case study in path analysis. *Journal of Educational Statistics, 12*, 101–128.
Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association, 81*, 945–960.
Hope, K. (1984). *As others see us: Schooling and social mobility in Scotland and the United States.* New York, NY: Cambridge University Press.
James, L. R., & Brett, J. M. (1984). Mediators, moderators, and tests for mediation. *Journal of Applied Psychology, 69*, 307–321.
Kenny, D. A. (1979). *Correlation & causality.* New York, NY: Wiley.
Liu, K. (1988). Measurement error and its impact on partial correlation and multiple linear regression analysis. *American Journal of Epidemiology, 127*, 864–874.
MacKinnon, D. P., Krull, J. L., & Lockwood, C. M. (2000). Equivalence of the mediation, confounding and suppression effect. *Prevention Science, 4*, 173–181.
Maruyama, G. (1998). *Basics of structural equation modeling.* London, England: Sage.
McClendon, M. J. (1994). *Multiple regression and causal analysis.* Itasca, IL: Peacock.
Meehl, P. E. (1978). Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. *Journal of Consulting and Clinical Psychology, 46*, 806–834.
Mulaik, S. A. (1987). Toward a conception of causality applicable to experimentation and causal modeling. *Child Development, 58*, 18–32.
Nickerson, R. S. (2000). Null hypothesis significance testing: A review of an old and continuing controversy. *Psychological Methods, 5*, 241–302.
Pedhazur, E. J. (1997). *Multiple regression in behavioral research: Explanation and prediction* (3rd ed.). New York, NY: Wadsworth.
Rogosa, D. (1987). Causal models do not support scientific conclusions: A comment in support of Freedman. *Journal of Educational Statistics, 12*, 185–195.
Rozeboom, W. W. (1960). The fallacy of the null-hypothesis significance test. *Psychological Bulletin, 57*, 416–428.

Rozeboom, W. W. (1997). Good science is abductive, not hypothetico-deductive. In L. Harlow, S. A. Mulaik, & J. H. Steiger (Eds.), *What if there were no significance tests?* (pp. 335–391). Mahwah, NJ: Erlbaum.
Schmidt, F. L. (1996). Statistical significance testing and cumulative knowledge in psychology: Implications for the training of researchers. *Psychological Methods*, *1*, 115–129.
Schmidt, F. L., & Hunter, J. E. (1997). Eight objections to the discontinuation of significance testing in the analysis of research data. In L. Harlow, S. A. Mulaik, & J. H. Steiger (Eds.), *What if there were no significance tests?* (pp. 37–64). Mahwah, NJ: Erlbaum.
Schmidt, F., & Hunter, J. E. (2002). Are there benefits from NHST? *American Psychologist*, *57*, 65–71.
Trafimow, D. (2003). Hypothesis testing and theory evaluation at the boundaries: Surprising insights from Bayes's Theorem. *Psychological Review*, *110*, 526–535.
Trafimow, D. (2006a). Multiplicative invalidity and its application to complex correlational models. *Genetic, Social, and General Psychology Monographs*, *132*, 215–239.
Trafimow, D. (2006b). Using epistemic ratios to evaluate hypotheses: An imprecision penalty for imprecise hypotheses. *Genetic, Social, and General Psychology Monographs*, *132*, 431–462.
Trafimow, D., & Rice, S. (2009). A test of the NHSTP correlation argument. *Journal of General Psychology*, *136*, 261–269.

Original manuscript received January 30, 2011
Final version accepted April 26, 2011

The Journal of General Psychology, 2011, *138*(3), 215–228

Object Naming in Dyslexic Children: More Than a Phonological Deficit

SUSANA ARAÚJO
LUÍS FAÍSCA
INÊS BRAMÃO
FILOMENA INÁCIO
Universidade do Algarve

KARL MAGNUS PETERSSON
Universidade do Algarve
Karolinska Institutet
Max Planck Institute
Donders Institute for Brain, Cognition and Behavior

ALEXANDRA REIS
Universidade do Algarve

ABSTRACT. In the present study, the authors investigate how some visual factors related to early stages of visual-object naming modulate naming performance in dyslexia. The performance of dyslexic children was compared with 2 control groups—normal readers matched for age and normal readers matched for reading level—while performing a discrete naming task in which color and dimensionality of the visually presented objects were manipulated. The results showed that 2-dimensional naming performance improved for color representations in control readers but not in dyslexics. In contrast to control readers, dyslexics were also insensitive to the stimulus's dimensionality. These findings are unlikely to be explained by a phonological processing problem related to phonological access or retrieval but suggest that dyslexics have a lower capacity for coding and decoding visual surface features of 2-dimensional representations or problems with the integration of visual information stored in long-term memory.

This work was supported by Fundação para a Ciência e Tecnologia (SFRH/BD/28488/2006, PTDC/PSI/64920/2006, PTDC/PSI-PCO/110734/2009, IBB/CBME, LA, FEDER/POCI 2010), the Max Planck Institute for Psycholinguistics, the Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, and the Swedish Dyslexia Foundation.
Address correspondence to Alexandra Reis, Departamento de Psicologia, Faculdade de Ciências Humanas e Sociais, Campus de Gambelas, Universidade do Algarve, 8005-139 Faro, Portugal; aireis@ualg.pt (e-mail).