Structural Equations Modeling

VI.A. STRUCTURAL EQUATIONS MODELING AND STATEMENTS REGARDING CAUSALITY

My question deals with an issue related to causation. How does one resolve issues of causal direction? For example, suppose I develop the model depicted in Figure 1, wherein materialism mediates the relation between "perceptions of marketing" and "life satisfaction."

As proposed, perceptions of marketing influence, or *cause,* materialistic values. A reviewer, however, suggests that the relation could just as easily be reversed; that is, materialistic values may cause a person to have favorable perceptions of marketing. Although there is some theoretical precedent for the proposed direction (e.g., Belk, 1985; Belk, Mayer, & Driscoll, 1984), the theory in this area is not well developed. Outside of a compelling theory, how can I defend the proposed direction? I have seen this problem arise in a number of situations (e.g., doctoral defenses, presentations), but I have not seen it satisfactorily resolved.

REFERENCES

Belk, Russell W. (1985). Materialism: Trait aspects of living in the material world. *Journal of Consumer Research, 12,* 265–280.
Belk, Russell W., Mayer, Robert, & Driscoll, Amy. (1984). Children's recognition of consumption symbolism in children's products. *Journal of Consumer Research, 10,* 386–397.

Professor Richard Netemeyer
Louisiana State University

This question addresses the issue of "mediation" as well as causal direction. I offer an answer to the causal direction por-
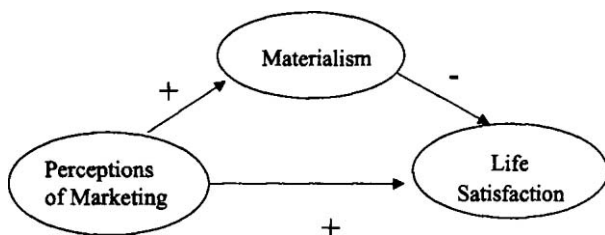

FIGURE 1   Structural equations attempting to capture causality.

tion of this question assuming nonexperimental data and the use of correlational data within a structural equation modeling (SEM) framework. As such, my answer is very narrowly focused on one type of methodology relative to assessing causality. The interested reader is strongly urged to consult several sources that discuss the process of mediation as well (e.g., Baron & Kenny, 1986; Brown, 1997; Cohen & Cohen, 1983; Holmbeck, 1997). (The issue of detecting mediation is treated in detail in Question VI.E. in this special issue.)

In a correlational framework, particularly where the data were collected cross sectionally and modeled using SEM, it is next to impossible to confidently infer that one construct causes another. The requirement that a construct must be shown to temporally precede its effect just does not exist. Furthermore, in many models, switching the direction between two constructs will not alter the fit of the model and may not alter the parameter estimate of association between the two constructs (Hayduk, 1996; Hoyle & Panter, 1995; MacCallum, Wegener, Uchino, & Fabrigar, 1993; Mueller, 1997). Models such as these have been termed *equivalent models* because they comprise the same constructs as a specified target model that are no more or no less parsimonious than that target model. As noted by MacCallum et al., equivalent models raise a problem similar to a confound in an experimental design in that a confound offers an alternative explanation to the experimental findings. An equivalent model in SEM cannot be ruled out as a statistically equivalent representation of the data.

Furthermore, the conditions for establishing causality using SEM are the same for other techniques—independent variables (IVs) must be isolated, association must be demonstrated, and directionality must be established. Bollen (1989) offered a detailed review, and Hoyle and Smith (1994, pp. 438–439) offered a concise review of these conditions with respect to SEM. With cross-sectional SEM data, one can safely argue that a model is consistent with a set of causal relations, but the exact directional association cannot be demonstrated. Still, the determination of directionality is not solely a statistical judgment. Directionality judgments can be enhanced by logical reasoning and a thorough understanding of accumulated theory and research such that a compelling directionality argument can be made. In this case, a cause and effect sequence may be proposed but not fully tested using SEM.

Though still not adequate for a confident causal inference, longitudinal SEM procedures offer a better possibility of initial evidence for the direction of causation (Bullock, Harlow,

& Mulaik, 1994). There are many long-term studies that have used such an approach that provide stronger evidence for a temporal sequence of constructs in a model. See Bullock et al. (p. 256) and Hoyle and Smith (1994, p. 439) for a brief review of some of these applications.

In sum, resolving the issue of causal directionality with correlational data in a SEM framework is extremely difficult. As stated by many modeling researchers, causal inferences from such models (particularly those with cross-sectional data) are rare and are likely ill advised (Bullock et al., 1994; Hoyle & Smith, 1994; MacCallum et al., 1993). In many cases, the conditions for inferring causality will not be met or equivalent models that offer an alternative ordering to the model constructs that fit the data equally well will exist. Though putative logic, strong and learned theoretical arguments, and longitudinally collected data can clearly strengthen a directionality argument, most SEM applications are best viewed as potential explanations that are consistent with a set of causal relations. Ultimately, it is the design, not the statistical method (i.e., SEM), that permits causal hypotheses to be adequately tested (Bullock et al., 1994; Campbell & Stanley, 1963; Hoyle & Smith, 1994; Sobel, 1993).

REFERENCES

Baron, Reuben M., & Kenny, David A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51,* 1172–1182.

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

Brown, Roger L. (1997). Assessing specific mediational effects in complex theoretical models. *Structural Equation Modeling, 4,* 142–156.

Bullock, Heather R., Harlow, Lisa L., & Mulaik, Stanley A. (1994). Causation issues in structural equation modeling research. *Structural Equation Modeling, 1,* 253–267.

Campbell, Donald T., & Stanley, Julian C. (1963). *Experimental and quasi-experimental designs for research.* Chicago: Rand McNally.

Cohen, Jacob, & Cohen, Patricia. (1983). *Applied multiple regress/correlation for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Hayduk, Leslie. (1996). *LISREL issues, debates, and strategies.* Baltimore: Johns Hopkins University Press.

Holmbeck, Grayson N. (1997). Toward terminological, conceptual, and statistical clarity in the study of mediators and moderators: Examples child-clinical and pediatric psychology literatures. *Journal of Consulting and Clinical Psychology, 65,* 599–610.

Hoyle, Rick H., & Panter, Abigail. (1995). Writing about structural equation models. In Rick H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 158–176). Thousand Oaks, CA: Sage.

Hoyle, Rick H., & Smith, Gregory T. (1994). Formulating clinical research hypotheses as structural equation models: A conceptual overview. *Journal of Consulting and Clinical Psychology, 62,* 429–440.

MacCallum, Robert C., Wegener, Dwayne T., Uchino, Bert N., & Fabrigar, Leandre R. (1993). The problem of equivalent models in applications of covariance structure analysis. *Psychological Bulletin, 114,* 185–199.

Mueller, Ralph O. (1997). Structural equation modeling: Back to basics. *Structural Equation Modeling, 4,* 353–369.

Sobel, Michael E. (1993). Causal inferences in latent variable models. In Kenneth Bollen & John Long (Eds.), *Testing structural equation models* (pp. 3–35). Newbury Park, CA: Sage.

Professor Peter Bentler
University of California, Los Angeles

Your question is an excellent one, but the answer is difficult. The ideal design would be one in which causation across time can help eliminate one of the alternatives. In your example, if perceptions of marketing were measured at Time 1, materialism at Time 2, and life satisfaction at Time 3, it would be hard to argue that materialism at Time 2 affects perceptions at Time 1. Of course, in the typical cross-sectional design, these alternatives cannot be teased apart in this way. In such a design, the problem is that the two alternative models are not nested models, so that they cannot be directly compared statistically. Indeed, each of the two proposals is a saturated model, and so both will exactly reproduce any data set. This creates the quandary you suggested.

I do have a suggestion for a way to resolve this problem. One way to resolve this issue is to nest both models inside a larger model. Instead of a one-way arrow (in either direction) between perceptions and materialism, replace this with two one-way arrows in which perceptions affects materialism and vice versa. Now this becomes a nonrecursive model, which is not identified because it contains one more parameter than data points. Therefore, by itself it cannot be tested. However, if you add another variable or set of variables to the system that affects either materialism or perceptions, but not any of the other variables in the system, you will have an identified model. An example might be socioeconomic status, which one might propose to affect materialism but not any of the other variables. Or, you could add a variable or set of variables consequent to perceptions or marketing, but not both. If such a model is identified, your alternative hypotheses have to do with the relative size of the two paths between perceptions and materialism. For example, you may find that the path perceptions → materialism is significantly larger than the reverse path materialism → perceptions, which would support your initial ideas. As far as I know, this proposed solution to your problem has not appeared previously in the literature. A good basic discussion of nonrecursive models is given by Kline (1998). For example, you could embed your model into a diagram as given by his Figure 6.5.

REFERENCE

Kline, Rex B. (1998). *Principles and practice of structural equation modeling.* New York: Guilford.

Editor: As Bentler indicates, part of the problem with the question posed is that the researcher is working with a sim-

plistic, impoverished, nomological network. Adding paths representing influences from unique antecedents would indeed help tease apart the causal direction (as would adding distinct consequential variables). The proposed model does not have the usual conceptual limitation of a poor nomological net, which is the structural equivalence of two (or more) pairs (or sets) of nodes or constructs (i.e., identical precursors and effects). However, either scenario could be salvaged somewhat with greater theoretical complexity, so as to rule out alternative explanations, here, in trying to determine which causal direction might be more sensible.

Cliff (1983) presented a readable, sobering reminder of issues regarding making causal claims on the basis of modeling covariance matrices. For example, he stated, "data do not confirm a model, they only fail to disconfirm it" (p. 117). It is odd indeed that we (post modernists) claim to be philosophically inclined toward the orientation of falsification, yet articles are written more often from a perspective as if we seek to confirm that our models fit (in structural equations, or even hypotheses posed in an analysis of variance [ANOVA] framework). Cliff goes on to describe a "corollary that when the data do not disconfirm a model, there are many other models that are not disconfirmed either" (p. 117). Presumably, scientific progress is impeded with the existence of a single alternative competing model; MacCallum et al. (1993, cited earlier) demonstrated on models published in prestigious journals that there were astronomical numbers of alternatives models that would have provided equivalent fit.

There are all kinds of circumstances that might lead us to investigate correlational data (and Hulland, Chow, & Lam, 1996, documented the rise in use of SEM in marketing), but if we seek to make causal statements, we would best operate experimentally. "The most satisfactory, almost the only satisfactory, method for demonstrating causality is the active control of variables" (Cliff, 1983, p. 119).

REFERENCES

Cliff, Norman. (1983). Some cautions concerning the application of causal modeling methods. *Multivariate Behavioral Research, 18,* 115–126.
Hulland, John, Chow, Yiu Ho, & Lam, Shunyin. (1996). Use of causal models in marketing research: A review. *International Journal of Research in Marketing, 13,* 181–197.

VI.B. SEM USING CORRELATION OR COVARIANCE MATRICES

Why, when doing a SEM, are we supposed to use covariance matrices, rather than correlation matrices, even though a correlation matrix is a special kind of covariance matrix?

Professor Peter Bentler
University of California, Los Angeles

The most basic answer is that in the standard model setup, you should not use the correlation matrix primarily because of the laziness of the program developers such as myself. The fact is that the statistical theory for analyses based on the correlation matrix was developed some decades ago, but developers have not generally made it available in popular programs.

A more complicated answer is that, in a variety of contexts, it does not make sense to analyze correlations, because variance differences may be critical to understanding a phenomenon. For example, in repeated measures such as growth curve modeling, one may expect the variances of the variables (or factors) to increase systematically with time. Or, in multiple group models, differences in means and variances may be fundamental to understanding the group differences. Therefore, if a latent factor of vocabulary skills is measured by some vocabulary tests at several ages, it may be that in each case a one-factor model is appropriate, but only if the means and variances of the factor are allowed to increase with age. In such cases, analysis of correlation matrices would yield parameters that distort the phenomenon. Some phenomena also are simply best interpreted in the covariance or raw score metric. Suppose that part of a model involves the regression of annual income on years of schooling. Now, the unstandardized beta is important because it represents the expected increase in (say, dollar) income per year of added schooling, holding everything else constant. However, standardization forced through the correlation matrix makes it impossible to get this interpretation.

Of course, a covariance structure model also has a standardized solution. In this solution, all the parameters are interpreted as if all variables, including the data, are in a correlation metric. The advantage is that beta weights are now interpretable as standardized betas, which will make more sense than unstandardized betas when the variables have only arbitrary scales. Therefore, you get the advantage of the correlation metric even if you use a covariance matrix. Finally, in a single group, you could input the correlation matrix but treat it like any other covariance matrix (i.e., paying no special attention to the fact that variances happen to be 1.0). In that case, the statistics typically would be meaningful.

Professor Richard Bagozzi
University of Michigan

This is a complicated question and might be best answered in pieces. There are at least two cases in which a covariance matrix should be used and not a correlation matrix. One is when multiple sample analyses are done and constraints are placed on parameters across samples to test (by use of chi-square difference tests) for invariance of focal parameters. A second is

when, even when dealing with a single sample, equality constraints are imposed on two or more parameters (e.g., sometimes factor loadings are constrained to be equal or error variances are constrained to be equal). In both of these cases, the unique metric of measures for each sample are important to retain and take into account, and use of a correlation matrix, which imposes a kind of standardization, should be avoided.

More generally, however, the theory behind the maximum likelihood method is based on the covariance matrix, and, strictly speaking, the chi-square test and standard errors (SEs) of parameter estimates are not correct when the correlation matrix is used. This applies in general, even when a single sample is investigated and no equality constraints are imposed on parameters. A nice discussion of the issues can be found in Cudeck (1989).

Two further comments follow: Cudeck (1989) acknowledged that for single sample confirmatory factor analyses in which no equality constraints on parameters are imposed, the correlation matrix can be used for input. At least the parameter estimates have desirable properties, even if the chi-square test and SEs are not correct, strictly speaking. In practice, the chi-square test usually does not differ, if one uses the covariance or correlation matrix as input. As a practical analysis strategy, at least in early stages of analysis, it is useful to work first with a correlation matrix, as the parameters are bounded nicely and it is easier to make diagnoses of poorly fitted models and other problems by examining output produced by correlation matrices. With use of covariance matrices and depending on the complexity of a model, the difference in variance of measures (e.g., some measures may have very small variances, others very large variances, etc.), it is sometimes difficult to check for the sources of problems, as factor loadings, error variances, and so on, are all over the map and it is difficult to visually discover problems in convergent and discriminant validity by inspection of a covariance matrix.

Some further comments follow: For weighted least squares analysis of certain models in which, for example, intercepts are estimated (e.g., in models with multiplicative indicator variables), it is useful to analyze the augmented moment matrix in which means, a covariance matrix, and asymptotic covariance matrices are computed in certain estimation procedures.

REFERENCE

Cudeck, Robert. (1989). Analysis of correlation matrices using covariance structure models. *Psychological Bulletin, 105,* 317-327.

Professor Robert Cudeck
University of Minnesota

The statistical theory underlying SEM has mostly been developed assuming the analysis applies to covariance matrices.

On the other hand, most analyses actually conducted are performed with sample correlation matrices because correlations are easier to work with, and easier to understand, than covariances. Moving from variances and covariances to correlations involves a rescaling of the variables that has to be accommodated in the (statistical) theory or the results may be adversely affected. Most topics dealing with correlations in statistics are much more complicated than the related topics that treat covariances. It is fortunate that estimates and tests obtained in SEMs are scale free in the majority of cases. The practical result of this property is that in most applications, an analysis may be based on correlations. The results one obtains will be directly related to the corresponding results computed from variances and covariances. Cudeck (1989) gave an overview of the issues.

REFERENCE

Cudeck, Robert. (1989). Analysis of correlation matrices using covariance structure models. *Psychological Bulletin, 105,* 317-327.

Editor: As these experts indicate, there are some conditions under which a correlation matrix may be modeled and other conditions under which the default choice, the covariance matrix, is still preferred. Cudeck (1989) presented the issues in terms of whether a model is scale invariant (i.e., the results estimated on a correlation matrix would be simply a rescaled version of what one would obtain had the covariance matrix been modeled instead), and whether a parameter within a model is scale free (defined similarly). Two common scenarios of factor analysis are scale invariant (Cudeck, 1989, p. 318)—for example, $\Lambda' = \begin{pmatrix} \lambda_1 & \lambda_2 & 0 & 0 \\ 0 & 0 & \lambda_3 & \lambda_4 \end{pmatrix}$ with phi, the matrix of factor intercorrelations, scaled as a correlation matrix, and $\Lambda' = \begin{pmatrix} 1 & \lambda_1 & 0 & 0 \\ 0 & 0 & 1 & \lambda_2 \end{pmatrix}$ with phi as a covariance matrix. These invariant scalings mean that although the factor loadings and other parameters will not necessarily be identical for the covariance and correlation solutions, one may be rescaled to obtain and the other. In a sense, then, the results are "correlated" in that relatively large (or small) loadings in the covariance solution will correspond to the relatively large (or small) loadings in the correlation solution.

Cudeck (1989) did warn, however, of the problems of invariance, with the result that the factor patterns, the overall fit statistics, and the SEs of the parameter estimates (and therefore their significance tests) are likely to differ (across the analyses of a covariance and correlation matrix on the same data). He offers several tests, none of which are sufficient to demonstrate that the correlation matrix may be preferred but are necessary in that the violations of which would suggest the covariance matrix would be preferred (e.g.,

check that the diagonal of the model-produced predicted matrix for a correlation matrix is indeed comprised of unities, p. 320; fit the factor pattern model to both the correlation and covariance matrices and verify that the fit statistics are identical, p. 322). He also warned that most software has been programmed to produce SEs as if the modeled matrix were a covariance matrix, so it is his pessimistic and scary hunch that most SEs (and therefore significance tests) reported in the journal articles that model correlation matrices are probably wrong.

REFERENCE

Cudeck, Robert. (1989). Analysis of correlation matrices using covariance structure models. *Psychological Bulletin, 105*, 317–327.

VI.C. IMPROVING MODEL FIT BY CORRELATING ERRORS

In SEMs, the fit of a model can often be significantly improved by explicitly allowing for additional error covariances between manifest variables in the measurement model. What does this imply for interpretation of the resulting coefficients?

Professor Joseph Cote
Washington State University

It is widely known that SEMs are sensitive to model misspecification. In other words, changing the specification of the model will significantly affect the parameter estimates. As such, researchers must provide strong theoretical justification for any parameter specifications. Otherwise, the change in model specification may simply be capitalizing on chance, making the parameter estimates meaningless.

The need for strong theoretical justification applies to all structural and measurement parameters, not just correlated errors. Although reviewers commonly do not question changes in structural parameters (e.g., allowing correlation among exogenous variables, freeing paths between endogenous and exogenous variables, etc.), this does not relieve the researcher from the burden of providing strong theoretical justification. Conversely, reviewers often raise concerns about freeing correlated error parameters. As noted earlier, their concerns are perfectly justified.

It is also true that strong theoretical justification can be provided for specifying a correlated error (e.g., method effects can be modeled as a latent construct or correlated errors). As long as all model specifications (including correlated errors) can be theoretically justified, then interpretation of the parameter estimates is appropriate. If any model specification (including structural relations) can-

not be theoretically justified, the interpretation of parameter estimates is inappropriate.

This question raises a second concern that is not explicitly stated. What constitutes a significant fit? The use of the chi-square test for model fit has well-know limitations. Researchers will commonly use one of several fit indexes applying a rule of thumb as a guide (e.g., normed fit index greater than 0.90). There exists another severe, but less well-known, problem with fit indexes. When there are a large number of measures or constructs, it is not uncommon for fit to be degraded. It may well be that no amount of model modification will yield an acceptable fit using standard approaches. In cases in which models are very complicated (i.e., large number of measures and constructs), Carmines and McIver (1981) suggested the following:

> Wheaton et al. (1977) suggest that the researcher also compute a relative chi-squared (c^2/df). As they indicate, this statistic takes sample size into consideration in assessing goodness of fit. They suggest a ratio of approximately five or less, "as beginning to be reasonable." In our experience, however, chi-square to degrees of freedom ratios in the range of 2 to 1 or 3 to 1 are indicative of an acceptable fit between the hypothesized model and the sample data. (p. 80)

My experience indicates that complex models with fit indexes in the 0.8 range would still meet the acceptability requirements suggested by Carmines and McIver (1981). I would rather see an author make this argument rather than use post hoc model modifications to get a normed fit index greater than 0.9. However, if a reviewer or editor insisted on a normed fit index greater than 0.9, then I would consider a post hoc correlated error if the author could demonstrate that all other parameter estimates were unaffected by the change. The stability of the estimates would provide an indication that the model was not purely capitalizing on chance.

REFERENCE

Carmines, Edward G., & McIver, John P. (1981). Unobserved variables. In George W. Bohrnstedt & Edgar F. Borgatta (Eds.), *Social measurement: Current issues* (pp. 111–130). Beverly Hills, CA: Sage.

Professor Richard Netemeyer
Louisiana State University

I restrict my answer to this question to cases involving within-factor correlated measurement error. First, most applications of allowing for within-factor correlated measurement errors (i.e., additional error covariances between manifest variables) to improve model fit represent post hoc adjustments (Anderson & Gerbing, 1988; Bagozzi, 1983; Fornell, 1983; Gerbing & Anderson, 1984; Hoyle & Smith,

1994; MacCallum, Roznowski, & Necowitz, 1992). Thus, the likelihood is high that this correlated error is sample idiosyncratic and may not replicate to other samples. Second, although allowing for a correlated measurement error will almost always improve fit, the researcher still may know very little as to why such an error is present in the first place—the reasoning for such error is likely speculative. Third, measurement models that contain within-factor correlated measurement errors do not represent unidimensional constructs. Some unwanted or unexplained sources of variance exist that were not specified a priori in the model. That is, the covariation between a pair of indicators (items) has not been adequately accounted for by the factors in the original measurement model. This unwanted or unexplained covariation could be due to many things (i.e., a methods factor, respondent yea-saying, the presence of an unwanted theoretical trait, etc.), but it still threatens unidimensional measurement. Thus, by allowing for within-factor correlated measurement errors, the correspondence between the construct of interest and its empirically determined factor becomes unclear (Gerbing & Anderson, 1984).

In sum, the use of within-factor correlated measurement error to improve model fit should be viewed critically. There will be few occasions where such error terms will be justified a priori and theoretically. In fact, several researchers advocate that a within-factor correlated measurement error should not be employed unless (a) it is warranted on theoretical or methodological grounds (e.g., correlating measurement errors for the same indicator for the same constructs for longitudinally collected data), (b) it does not significantly alter the structural parameter estimates of a model, and (c) it does not significantly alter the measurement parameter estimates of a model (Bagozzi, 1983, p. 450; Fornell, 1983, p. 447; Gerbing & Anderson, 1984, p. 579). Gerbing and Anderson further suggested that the strongest emphasis be placed on the first condition.

Professor Peter Bentler
University of California, Los Angeles

Implied in your question is the simple answer—namely, that there are additional sources of correlation in the variables beyond only the common factors that are part of the measurement model. Stated differently, the model has omitted some common factors, where each correlated error represents an omitted factor. (It will be difficult to identify such factors with only two indicators per factor, so the correlated error representation is a lot more reliable, as well as simpler to obtain, based on the Lagrange Multiplier test, as I call it in my program EQS, or on modification indexes.)

It seems reasonable to be suspicious of such correlated errors primarily because of their post hoc status, which might imply that they would not cross validate well, especially when initially obtained from a small sample. (I do not worry much if they are obtained in very large samples.) Aside from this reasonable criticism, worry about correlated errors seems to me to be overblown. It is very likely that in any domain, theory will always be inadequate to exactly predict all sources of correlation in data. For example, when building an instrument with 20 items to measure a single factor, it may well be that a very large general factor could largely explain the correlations among the items. However, it probably would be foolhardy to expect that there are not minor sources of covariation due to such things as common item wordings, item sequences, unanticipated minor factors, and perhaps violation of continuity assumptions through the use of Likert scales. Correlated errors will summarize such covariation. Do not forget, however, that current methods are not good enough to construct, post hoc, any omitted common factors that might exist in such error correlations. If there are many correlated errors, it may be best to go back to exploratory factor analysis to check on the main latent sources of variance to clean up the measurement model.

REFERENCES

Anderson, James C., & Gerbing, David W. (1988). Structural equation modeling in practice: A review and recommended two-Step approach. *Psychological Bulletin, 103*, 411–423.

Bagozzi, Richard P. (1983). Issues in the application of covariance structure analysis: A further comment. *Journal of Consumer Research, 9,* 449–450.

Fornell, Claes. (1983). Issues in the application of covariance structure analysis: A comment. *Journal of Consumer Research, 9,* 443–448.

Gerbing, David W., & Anderson, James C. (1984). On the meaning of within-factor correlated measurement errors. *Journal of Consumer Research, 11,* 572–580.

Hoyle, Rick H., & Smith, Gregory T. (1994). Formulating clinical research hypotheses as structural equation models: A conceptual overview. *Journal of Consulting and Clinical Psychology, 62,* 429–440.

MacCallum, Robert C., Roznowski, Mary, & Necowitz, Lawrence B. (1992). Model modifications in covariance structure analysis: The problem of capitalization on chance. *Psychological Bulletin, 11,* 490–504.

VI.D. SEM VERSUS ANOVA

A researcher proposes a conceptual model representing hypothesized causal relations among a set of antecedent and outcome constructs. An experimental study is undertaken in which certain antecedents are both manipulated and measured. Outcomes are also measured. ANOVA on the manipulation checks reveals that the manipulations have the expected effect on participants' responses to the antecedent measures, thereby supporting the success of the manipulations. In contrast, ANOVAs on the outcome measures reveal that participants' responses to these are unaffected by the manipulations. However, when the researcher uses causal modeling analyses based on participants' responses to the antecedent and outcome measures, there are significant path

estimates between the antecedents and outcomes. What conclusions should the researcher draw about the hypothesized causal relations? Is it appropriate to interpret the correlational findings as supporting hypothesized causal relations, despite the null effects observed for the outcome measures in the experimental study?

Professor Joseph Cote
Washington State University

Oooh, this is a good one! First, you must have faith that the structural model is accurate. This is not something that should be quickly assumed. It is well known that SEM is very sensitive to model misspecification. As such, I would instantly have some doubts about the LISREL estimates. At the same time, it is possible that a true effect is not picked up by ANOVA. By default, ANOVA reduces the available information (converting interval data to categorical data). It is possible that ANOVA would miss a relation that would be uncovered by LISREL. To be honest, a researcher would need to look closely at the data to decide which approach was most appropriate. I would also expect this situation to arise only when the relations (explained variances) were relatively small (even if statistical significance is very strong). If you had a strong relation, the information reduction caused by ANOVA would not affect the results. I should also note that most researchers fail to report the variance explained when fitting LISREL models. They simply assume that because statistical significance exists, relations are strong. This is simply not true.

Professor Richard Bagozzi
University of Michigan

Actually, in addition to the ANOVA and causal modeling analysis, an analysis of the experimental data can be made as well by use of SEM in an ANOVA design framework, but where measurement error is taken into account. Doing so is especially useful for multivariate analysis of variance (MANOVA) contexts in which causal orderings are proposed also among the dependent variables (DVs) and a type of step-down analysis is to be conducted. A limitation of this approach in practice, however, is the need to have large samples.

The problem with a traditional causal model, as with path analysis, is that the findings of the significant paths between measured manipulations (e.g., manipulation checks taken on interval scales) and outcome measures indicate only a linear association between the variables. I am not sure, however, how often the proposed discrepancy occurs in practice between null effects under ANOVA and significant paths under causal models. It is important to get at the bottom of the discrepancy. One possibility is that so much error exists in the manipulation

as to inflate the denominator in the F test, but with multiple measures of manipulation checks, it is possible to correct for measurement error when doing the causal analysis.

Editor: Bagozzi's perspective on this issue is based on recent work done by him and other researchers who have considered a closer integration of the analysis of variance and SEMs (cf. Bagozzi & Yi, 1989; Cole, Maxwell, Arvey, & Salas, 1993). It is important in translating an ANOVA problem into a SEM framework, that the basics are not lost (e.g., most SEMs are only main effects driven models, perhaps due to the multicollinearity problems interaction terms induce; cf. Jaccard, Turrisi, & Wan, 1990, pp. 8, 30).

Bagozzi and Yi (1989) claimed the relative advantages of SEMs over MANOVA include no restrictive equality of covariance matrices across conditions, an opportunity to correct for measurement error, and a more complete modeling of the theoretical relations among all the variables (p. 282). The relative disadvantages include less robustness to possible violations of the assumption of multivariate normality, requirements of larger sample sizes, and more complex models with larger numbers of parameters to estimate increase in the likelihood of improper solutions and nonconvergence (p. 283).

There is another issue in the question that was posed—the question implies that the researcher has conceptualized a sort of process by which the manipulated variables are thought to impact most immediately the manipulation checks and then more distally the outcome measures, almost as if the manipulation check variables are mediators. Perhaps the experimental procedure lends itself to effects that are that short lived; possibly the manipulation needs to be stronger. (Its weakness may be indicated by effect size indexes or relevant path coefficients.)

REFERENCES

Bagozzi, Richard P., & Yi, Youjae. (1989). On the use of structural equation models in experimental designs. *Journal of Marketing Research, 26*, 271–284.
Cole, David A., Maxwell, Scott E., Arvey, Richard, & Salas, Eduardo. (1993). Multivariate group comparisons of variable systems: MANOVA and structural equation modeling. *Psychological Bulletin, 114*, 174–184.
Jaccard, James, Turrisi, Robert, & Wan, Choi K. (1990). *Interaction effects in multiple regression.* Newbury Park, CA: Sage.

VI.E. MEDIATORS AND MODERATORS

I have been looking into articles on mediating and moderating. The main citations appear to be Baron and Kenny (1986) and Sharma, Durand, and Gur-Arie (1981). They say that to

show that C mediates the relation between A and B ($A \to C \to B$), you have to model

1. $B = b_0 + b_1 A$, and show that A (the IV) has a significant main effect on B (the ultimate DV).
2. $C = b_2 + b_3 A$, and show that A has a significant main effect on C (the mediator).
3. $B = b_4 + b_5 A + b_6 C$, and show not only that C has a main effect on B, but also that b_5 (the b coefficient for A in Equation 3) is less than b_1 (the coefficient for A in Equation 1).

One thing that they do not make clear is whether, in a multiple regression framework, these tests should be done individually or as part of the whole multiple regression model. I think that the tests should be done in the context of the whole model, but because both articles deal with simple regression, it is not clear, and in fact gives the impression that the tests should be done separately and individually. For instance, suppose I am predicting E using A, B, and C, and that I think D mediates between B and E (i.e., $A \to E$, $B \to D \to E$, and $C \to E$). It seems to me that I would model

4. $E = b_7 + b_8 A + b_9 B + b_{10} C$, and show that, in the context of (i.e., while statistically controlling for) the other variables, B has a significant main effect on E.
5. $D = b_{11} + b_{12} B$, and show that, like in Equation 2, the focal IV has a significant effect on the mediator (or should this be, $E = b_0' + b_1' A + b_2' D + b_3' C$).
6. $E = b_{13} + b_{14} A + b_{15} B + b_{16} C + b_{17} D$, and show that, even in the context of the other variables, D has a significant main effect on E, and that b_{15} (for B here) is less than b_9 (B in Equation 4).

Anyway, if this is the right way to do it, I cannot find anyone who says clearly that it is. Which means it would be good to have it written down somewhere. And if it is the wrong way to do it, then it is even more important to set people like me straight.

Moderation issues appear just as tricky. Sharma et al. (1981) seem to be arguing for the following: To prove a moderating influence of C on A's relation with B, you must model

7. $B = b_{18} + b_{19} A$, to show a main effect for A.
8. $B = b_{20} + b_{21} A + b_{22} C$, and show no main effect for C because if there was a main effect it would be an IV or a "quasi moderator" but not a moderator.
9. $B = b_{23} + b_{24} A + b_{25} C + b_{26} A * C$, and show an effect for the $A * C$ interaction, and a lower beta coefficient for A in this equation (b_{24}) than (b_{19}) in Equation 7.

Therefore, they argued for inclusion of the main effect, which I think a lot of people do not do. Second, what they actually argued in their article is not described in the way I have put it—

this is my interpretation of their notation, so here again I think it is worth having a more definitive word.

While investigating these issues, I have come across two Web sites on mediation and moderation that readers may find useful: (a) David Kenny's site, http://nw3.nai.net/~dakenny/causalm.htm; and (b) David MacKinnon's site, http://www.public.asu.edu/~davidm/ripl/mediate.htm.

REFERENCES

Baron, Reuben M., & Kenny, David A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*, 1173–1182.

Sharma, Subash, Durand, Richard M., & Gur-Arie, Oded. (1981). Identification and analysis of moderator variables. *Journal of Marketing Research, 28*, 291–300.

Professor Donald Lehmann
Columbia University

Tests for mediation and moderation are widely employed, usually in a formulaic manner. However, these tests are only a means to the goal of developing a causal understanding of a phenomenon. Thus, before discussing such tests, three points seem relevant: (a) Both issues deal with proper specification of a simultaneous equation (structural) model. Correct specification involves a blend of theoretical and empirical issues. Whether one concludes that a particular relation exists depends on the strength of the priors, not just statistical significance in the sample. (b) Mediation and moderation are logically continuous constructs. Thus, the real question is about degree, not whether or not. (c) Although the term *moderate* implies a lessening of an effect, higher levels of one variable can lead to greater (or lower) effects of another. Thus, a better term would be *alter*.

The simplest way to consider mediation and moderation is in the context of simple graphical models (Figure 2). In Case A, the model posits that X affects Z. Model B adds a second predictor, Y. Model C assumes X affects Z only indirectly through its effect on Y (i.e., X's effect on Z is fully mediated by Y). Model D assumes the effect of Y on Z is affected or altered (i.e., moderated) by the level of W. Model E assumes both mediation (of Y on X) and moderation (of W on Y). Finally, Model F assumes the effect of X on Z is partially mediated by Y.

If the assumed model is simple until proven otherwise (the common scientific practice of parsimony even though the opposite may be true), then links are added only when they explain significant additional variation. That means we begin with the simple models and add terms. For example, to Model B, $Z = b_0 + b_1 Y + b_2 X + e$, we add b_3 (corresponding to the interaction term YW in Model D) to see if W moderates the effect of Y on Z. The test then becomes whether b_3 is significantly differ-

A. Simple Model: $Z = a_0 + a_1 X + e$

$X \longrightarrow Z$

B. Two Predictor Model: $Z = b_0 + b_1 Y + b_2 X + e$

$X, Y \longrightarrow Z$

C. Mediated Model: $Z = d_0 + d_1 Y + e$
$Y = c_0 + c_1 X + e$

$X \longrightarrow Y \longrightarrow Z$

D. Moderated (Interacted) Model:
$Z = b_0 + b_1 Y + b_2 X + b_3(YW) + e$

$X, Y \longrightarrow Z \quad W$

E. Mediated and Moderated Model:
$Y = c_0 + c_1 X + e$
$Z = b_0 + b_1 Y + b_2 X + b_3(YW) + e$

$X \longrightarrow Y \longrightarrow Z \quad W$

F. Partially Mediated Model:
$Y = c_0 + c_1 X$
$Z = b_0 + b_1 Y + b_2 X$

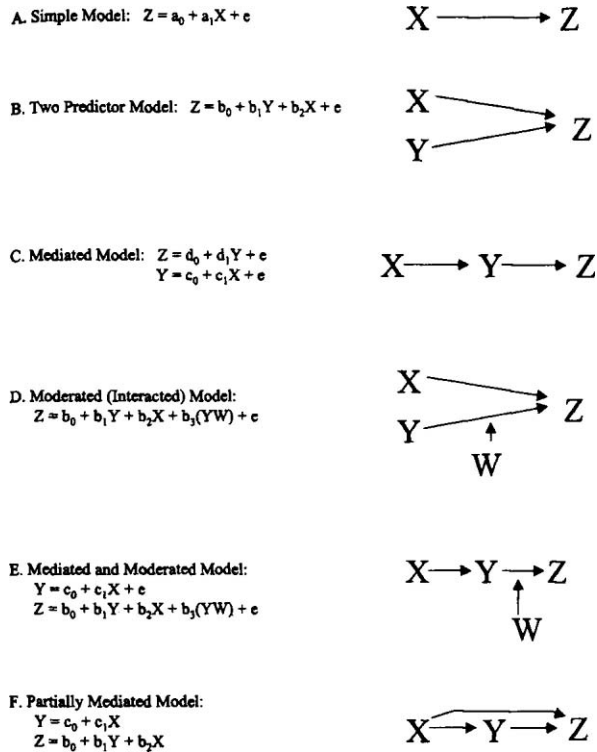$X \longrightarrow Y \longrightarrow Z$

FIGURE 2   Depicting different models to represent causal structures.

ent from zero, or equivalently, if adding the variable significantly increases the variance explained (i.e., R^2); that is, if Model D significantly outperforms Model B (e.g., Lehmann, Gupta, & Steckel, 1998, pp. 537–539). Notice this says nothing about how the size of the direct (main) effect of Y on Z changes; it can stay the same, increase, or decrease, which is largely determined by the coding scheme used.

Similarly, mediation requires two conditions. First, X must impact Y (i.e., $c_1 \neq 0$ in Model C). Second, the direct effect of X on Z decreases when we add Y to the model (i.e., b_2 in Model B is less than a_1 in Model A).[1] Thus, mediation requires that both direct and indirect (through Y) paths exist from X to Z and that the direct path in this model be weaker than the simple direct effect of X on Z (i.e., in Model A).

If there are a number of possible moderators (interactions), then two approaches are possible. First, one could add interactions one at a time and separately test for their significance. This makes sense if each has some theoretical support (i.e., a priori expectations of nonzero interactions). On the other hand, if we return to the simple until proven otherwise position, then some of the interactions may appear to be significant due to chance (e.g., 1 out of 20 at the .05 level). Therefore, a more conservative procedure is to test for the im-

provement due to the inclusion of the set of interactions as a group (i.e., in a nested model F test). Then, only if the set of interactions is significant does one examine individual terms.

It is important to notice that tests of both mediation and moderation are influenced by collinearity among predictor variables. In mediation tests, the original and mediator variables are correlated (in our example, X and Y), which makes estimation of the separate effects of X and Y less reliable (and hence, tests to prove they are nonzero more difficult). The problem is even worse for moderation tests. Whereas in mediation analyses we expect that X and Y should be correlated, the correlation between Y and YW is basically an inevitable inconvenience. (The problem gets worse if multiple interactions are tested simultaneously because Y, YW_1, YW_2, etc., all share a common element.) Therefore, various coding schemes (i.e., effect coding) are often employed to reduce the problem.

In running tests for mediation and moderation, also keep in mind that

1. Moderator analysis focuses on whether a particular known and measured variable influences the effect of one variable on another. When the moderator represents differences in participants, the issue is whether the effect is constant across all observations (e.g., participants). In many situations, heterogeneity in response is not well explained by the readily available variables (e.g., demographics). Thus, other means of exploring participant heterogeneity in response (e.g., cluster or latent class analysis) may be fruitful.

2. Treating moderation or mediation as a yes–no question leaves out a lot of information. Not only is statistical significance of $p = .051$ not very different from $p = .049$, but the p value itself is easily manipulated (i.e., by increasing sample size). Therefore, a more continuous measure that concentrates on the size of the effects seems preferable. For example, a simple index of mediation comparing a_1 in Model A to b_2 in Model F is $(a_1 - b_2)/a_1$. Assuming a_1 and b_2 are positive and b_2 is less than a_1, this index ranges from 0 (*complete mediation*) to 1 (*no mediation*).

3. Mediation tests derive directly from the principles of path analysis (Asher, 1976; Blalock, 1964). Specifically, we can compare a simple model of X to Z (Model A) to a two-equation model in which, in the first equation, X affects Y, and in the second, Y affects Z (Model C). If the effect of X to Z can be explained by the product of the effect of X on Y and Y on Z (i.e., $a_1 = c_1 d_1$), then there is no need for (nor evidence of) a direct link from X to Z. (In this simple case, that result is equivalent to having $r_{xz} = (r_{xy})(r_{yz})$.)

REFERENCES

Asher, Herbert B. (1976). *Causal modeling.* Beverley Hills, CA: Sage.
Baron, Reuben M., & Kenny, David A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51,* 1173–1182.

---

[1]For readers familiar with Baron and Kenny (1986), it may be easier to think of the b_2 in Model F as that which becomes insignificant, or at least less than a_1.

Blalock, Hubert M., Jr. (1964). *Causal inferences in non-experimental research*. Chapel Hill: University of North Carolina Press.

Lehmann, Donald R., Gupta, Sunil, & Steckel, Joel H. (1998). *Marketing research*. Reading, MA: Addison-Wesley-Longman.

Professor Roderick McDonald
University of Illinois

I think the best way to discuss the question of detecting or showing moderation or mediation in a regression is to set aside the entire literature on these topics and start from scratch. First, consider moderation and the regression equation for the interaction model:

$$Y = b_0 + b_1X + b_2Z + b_3XZ + e \qquad (1)$$

where Y is a (random) response variable, X and Z are explanatory variables—fixed or random, quantitative or categorical—and e is a residual. Standard texts such as Neter, Wasserman, and Kutner (1996) told us how to fit the regression, get SEs for the coefficients b_0, b_1, b_2, and b_3, and use these to get confidence bounds and test the parameters for significance—for those who, unlike me, still believe in significance tests (see Harlow, Mulaik, & Steiger, 1997). Not all texts on the linear model tell us this, but it is not necessary to distinguish cases in which the explanatory variables are random or fixed (when observed, they are no longer random), or, as Baron and Kenny (1986) do: to distinguish quantitative (continuous) and categorical IVs. If X and Y are highly correlated, this makes a problem for the naive computation of the SEs of b_1 and b_2, but if we fit the interaction model with X and Z measured from their means, then XZ is uncorrelated (by construction) with X and Z, so the SE of b_3 is trustworthy and there is no problem, except the unavoidable judgment, deciding whether it is significant and non-negligible—that is, whether we have a departure from additivity of the effects of X and Z.

If we chose to rewrite the regression as

$$Y = [b_0 + b_2Z] + [b_1 + b_3Z]X + e \qquad (2)$$

we might be tempted to say that the moderator variable Z alters the regression coefficient $[b_1 + b_3Z]$ of Y on the explanatory and IV X. If we chose to rewrite it as

$$Y = [b_0 + b_1X] + [b_2 + b_3X]Z + e \qquad (3)$$

we might be tempted to say that the moderator variable X alters the regression coefficient $[b_2 + b_3X]$ on the IV Z. As I see it, both temptations should be avoided, thus avoiding the whole collection of reifications and misconceptions that appear to constitute the literature on moderator variables. An interaction is a departure from additivity of effects that is symmetric in the explanatory variables.

There also seems to be some confusion in the literature you cite about moderating a regression and moderating a correlation. Using graduate school algebra of expectations, we easily find that the squared correlation between X and Y is given by

$$\rho_{X,Y}^2 = \frac{[b_1 + b_3Z]^2\sigma_X^2}{[b_1 + b_3Z]^2\sigma_X^2 + \sigma_e^2} \qquad (4)$$

and, symmetrically, the squared correlation between Z and Y is

$$\rho_{Z,Y}^2 = \frac{[b_2 + b_3Z]^2\sigma_Z^2}{[b_2 + b_3Z]^2\sigma_Z^2 + \sigma_e^2} \qquad (5)$$

so, we can say that if the effects are not additive, then each IV moderates the correlation of the DV with the other. However, why complicate the obvious? An interaction is an interaction is an interaction, as Gertrude Stein would have put it.

However, there is a complication deserving remark. Finding that one explanatory variable alters the correlation of the response with another does not necessarily imply the interaction model. It could be that in the simple model

$$Y = b_4 + b_5X + e \qquad (6)$$

the error variance of Y is a function of some variable Z. (This is only one of many models invoking control of heteroscedasticity or types of nonlinearity giving rise to a dependence of correlations on other measures.)

My recommendation is that, first, fit the model in Equation 1 to check for nonadditivity, measuring each variable from its mean. Do not call the nonadditivity moderation of either by the other. Second, do residual plots if you suspect heteroscedasticity caused by a variable included in your data and try to find a suitable model for it, but do not expect it as an everyday occurrence in research.

Now we turn to mediation. Again, I do not follow the conventional wisdom offered in the references you cite, although I do not reject the term. First, we recognize the following rather grim reality: It is a truism of path analysis, dating back to Sewell Wright (see Duncan, 1975, for a classic account) that if the partial correlation of X with Y when Z is partialled out is zero, then possibly (a) X directly causes Z and Z directly causes Y, but X has no direct effect on Y; (b) Y directly causes Z and Z directly causes X, but Y has no direct effect on X; and (c) Z directly causes both X and Y, and X and Y have no other common cause. We may feel we can rule out (c) on good theoretical grounds. I have seen few good examples of this, even when we have a clear time order, because the measures may reflect a developmental process or a state stable over time, so that "post" has nothing to do with "propter." If we can rule out

(c), we may say that Z mediates between X and Y in one direction or the other.

It is well known, though commonly not mentioned in the texts on SEM, that for model (a), separately fitting the two regressions (a) Z on X and (b) Y on Z, is equivalent to fitting the full model with SEM software by maximum likelihood (but not by least squares, curiously enough), and that testing the partial correlation ρ_{XYZ} is equivalent to testing fit of the full model. In fact, the simple regression and partial correlation procedure has small sample properties that the clever SEM approach lacks. Therefore, you can get the effect sizes by regression and test mediation—the channeling of causality through Z—by partial correlation, asking if the partial correlation is significant or, better, if it is approximately zero, just using Fisher's transformation as in the elementary accounts of testing correlations. Nothing fancy is needed.

However, the idea of looking for a mediation over and above a nonnegligible direct effect, as in the question, seems to me very dangerous—with or without theoretical elimination of model (c). It is a truism of path analysis (again, see Duncan, 1975) that the model in which X both directly causes Y and indirectly causes it through Z is just identified and unrestrictive, fitting every data set perfectly. It is not distinguishable by data from the very large number of other nonrestrictive models for three variables. If we know so very much that we can rule out all the other models, we can fit the model:

$$Z = b_6 + b_7X + e_Z \qquad (7)$$

and

$$Y = b_8 + b_9X + b_{10}Z + e_Y \qquad (8)$$

either by a SEM program or by the two regressions. The direct effect of X on Y is b_9. This is the conjectured change in Y if we could control X and increase it by one unit while holding Z constant. The total effect of X on Y through the causal chain $X \rightarrow Z \rightarrow Y$ (the conjectured change in Y if we could make X increase one unit and allow both Z and Y to change as a consequence) is $(b_9 + b_7b_{10})$. Some writers then define an indirect–mediated effect as the difference between the total effect and the direct effect of X on Y. I prefer not to, because this does not correspond to a reasonable thought experiment. (However, see Bollen, 1989, for a contrary view.) A super-confident investigator may not only do this, but compare the indirect effect (b_7b_{10}) with the direct effect b_9 and claim that the indirect effect—which can be negative, tending to cancel the direct effect—is in some sense non-negligibly greater than or less than the direct effect. This supposes more substantive knowledge than I find available in the path analyses that come to me as a consultant.

Note that there is no need to follow the advice you quote that we should also fit

$$Y = b_0' + b_1'X + e_Y$$

because by SEM theory we know that the total effect, $b_1' = b_9 + b_7b_{10}$, is already given by just Equations 7 and 8. (McDonald, 1998, chap. 17, gives an elementary account of these matters.) Note also that whether b_9 is less than or greater than b_1' depends on the sign pattern of b_9, b_7, and b_{10}, so I believe the advice you cite is simply incorrect. Perhaps the conclusion is that mediation in regression models should not be cut loose from the more general concepts of SEM. And note that the only thing tested in a path model is the vanishing of a handful of partial correlations and is almost always consistent with many plausible alternative models (MacCallum et al., 1993). Failure to recognize this fact is allowing the proliferation of a confident but weakly supported collection of empirical SEM studies.

REFERENCES

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

Duncan, Otis D. (1975). *Introduction to structural equation modeling.* New York: Academic.

Harlow, Lisa, Mulaik, Stanley A., & Steiger, James. (Eds.). (1997). *What if there were no significance tests?* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

MacCallum, Robert C., Wegener, Dwayne T., Uchino, Bert N., & Fabrigar, Leandre R. (1993). The problem of equivalent models in applications of covariance structure analysis. *Psychological Bulletin, 114,* 185–199.

McDonald, Roderick P. (1998). *Test theory: A unified treatment.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Neter, John, Wasserman, William, & Kutner, Michael H. (1996). *Applied linear statistical models.* Chicago: Irwin.

Professor Joseph Cote
Washington State University

The steps outlined for mediating effects are essentially correct. However, the outline suggests looking at the coefficients. This is improper. You must look at the change in R^2. The reason for this is that mediators may cause collinearity problems. The collinearity will affect the regression coefficients in unpredictable ways, making them inappropriate indicators of significance. Compare the change in R^2 in sequence. Fit the following three models (assume D moderates relations between E and B):

1. $E = \text{fn}(A, B, C)$
2. $E = \text{fn}(A, C, D)$
3. $E = \text{fn}(A, B, C, D)$

To test for mediating effects, the size of R^2 values should be as follows $3 \geq 2 > 1$. Note that the R^2 for Equation 3 can be equal to the R^2 for Equation 2 because D may completely moderate the effects of B. Of course, the empirical test is only relevant if a strong theoretical case can also be made.

The procedure outlined for testing moderating effects is essentially correct. However, as with mediators, you cannot use the beta coefficients to test them. With interaction effects you

definitely have collinearity problems. As such, the beta values are unstable in unpredictable ways. I would test for moderating effects in the following manner. First, fit the following two models (assumes D moderates the relation between E and B):

4. $E = fn(A, B, C)$
5. $E = fn(A, B, C, B*D)$

To test for moderating effects, the R^2 for Equation 5 is greater that Equation 4. It is perfectly acceptable to then fit a third equation.

6. $E = fn(A, B, C, D, B*D)$

D can still have a direct effect on E in addition to the moderating effect. In this case, R^2 for Equation 6 is greater that Equation 5. Note that because of collinearity problems, any of the coefficients from Equation 5 may now be lower or insignificant (including interaction effects). However, this does not mean that an interaction does not exist. Again, because of collinearity problems, the beta estimates are unstable and unpredictable. Only the R^2 remains unaffected by collinearity. It is also important that strong theory accompany tests for moderating effects. Because of collinearity, the interaction term may sufficiently model a direct effect of D.

Professor Timothy Heath
University of Pittsburgh

It is an additional challenge to test mediation and moderation in the context of other model terms. The inclusion of one predictor or set thereof affects tests of other predictors, so the choice of which predictors and terms to include in regression models is one of the more fundamental decisions a researcher can face. When making such decisions, sensitivity to at least three types of relations within predictors is required: The first is *mediation*, where one predictor causes another predictor, which then causes the DV. The second is *confounding*, where predictors covary for reasons other than predictor-to-predictor causation. And the third is *moderation*, where the effect of one predictor on the DV depends on (and varies across) levels of another predictor. I discuss mediation and confounding in response to a question on the former because both involve patterns of collinearity. I then address the moderation question that asks whether it is necessary to test a given interaction (product) term in the context of its constituent main effects. These questions appear to be related because each involves the effect of some model terms on tests of some target effects, but despite a surface similarity, mediation and moderation involve different statistical tests and vastly different underlying causal models (e.g., Heath, 1990).

In this response, I limit myself to major issues involved in simple ordinary least squares (OLS) applications. For in-depth treatments of path analyses and moderation tests in multiple regression, readers are encouraged to consult the fol-

lowing sources: applied regression and research texts devoted to causal assessments such as Cohen and Cohen (1983), Pedhazur (1982), Cook and Campbell (1979), and Judd and McClelland (1989); econometrics texts such as Johnston (1984), Judge, Griffiths, Hill, and Lee (1980), and Maddala (1977); causal modeling books such as Blalock (1971), Bagozzi (1980), and Davis (1985); and books addressing tests of moderation in multiple regression such as Aiken and West (1991) and Jaccard, Turisi, and Whan (1990).

As per the question, Baron and Kenny (1986) stated that to show that W mediates the relation between X and Y ($X \rightarrow W \rightarrow Y$), you have to demonstrate the following:

	Step 1:	$Y=fn(X)$	and show that X has a significant main effect on Y,
	Step 2:	$W=fn(X)$	and show that X has a significant main effect on W,
and	Step 3:	$Y=fn(X,W)$	and show not only that W has a main effect on Y, but also that X's coefficient drops significantly relative to Step 1 (preferably to nonsignificance).

If we are predicting Y using X and Z and we think W mediates between X and Y, should we include additional predictors (here, Z) in all of the submodel tests, or should we eliminate them?

In this example, X is our primary predictor, whose effect is proposed to go through W, our hypothesized (causal) mediator. Z is a confound or error control. We might include Z to (a) make sure that X does not receive undue credit for causing Z (i.e., an $X \rightarrow Y$ relation arising only because X is correlated with the actual cause, Z) or (b) to reduce error variance and thereby increase sensitivity to X's effects. As we will show, there is no statistical distinction between Z and W, which, by OLS, are treated essentially as covariates, but the interpretations of their effects diverge in important ways.

There are two general responses to this initial question. First, models that add a potential mediator (W) to assess its impact on the $X \rightarrow Y$ coefficient should retain all predictors from the original model testing the $X \rightarrow Y$ coefficient (e.g., Z) lest changes in X's coefficient arising from the omission of those variables be mistakenly attributed to the inclusion of the mediator. Second, the models used to test mediation depend on the researcher's a priori beliefs and goals. If there is reason to believe that the true effect of X on Y is that which exists after removing any potential effect of Z, then Z should be included in all models.

If, on the other hand, the researcher is not sure what pattern of variances best reflects X's effect on Y, the safest tack is to include Z in all models because inclusion would be consistent with conventional wisdom and popular path analytic approaches. The rationale is that the best specification is the one that includes all potentially relevant variables to (a) maximize statistical power by reducing error and (b) account as thoroughly as possible for potential collinear or confounding relations. In practice, therefore, researchers commonly test

mediation with packages such as LISREL and EQS in which numerous facets of the underlying measurement and structural models are estimated simultaneously. Hence, there is ample reason for adapting Steps 1 through 3 to broader-based models in the following manner, where Z can be thought of as either a single predictor or set thereof:

	Step 1':	$Y=\text{fn}(X,Z)$	to show that when extracting the variance in X and Y that is shared with other variables (Z), X and Y remain associated significantly,
	Step 2':	$W=\text{fn}(X,Z)$	to show that when extracting the variance in X and W that is shared with other variables (Z), X and W remain associated significantly,
and	Step 3':	$Y=\text{fn}(X,W,Z)$	to show that when extracting the variance in X, Y, and W that is shared with other variables (Z), not only that W is significantly associated with Y, but that the $X{\rightarrow}Y$ association becomes significantly weaker relative to Step 1' (preferably becoming nonsignificant).

However, in these cases in which researchers are not certain which patterns of shared variance with Y best reflect the causal influences of X, W, and Z, researchers may wish to report more thorough mediation tests. The reason is that despite the justification for including Z in all models, inclusion risks inflating the ambiguity already inherent in the statistical controls that anchor most tests of mediation. Consider the following patterns and interpretations when we control for Z (as a potential confound) either experimentally or statistically.

Nature of Means of Control for Z

Evidence of $X{\rightarrow}Y$ relationship	Experimentally Controlling Z (Manipulate X and Z Orthogonally)	Statistically Controlling for Zs Effects (Measure Z and Extract the Variance It Shares With X and Y Before Estimating the $X{\rightarrow}Y$ Association)
If evidence of X's effect (e.g., a significant $X{\rightarrow}Y$ association) *persists* when controlling for Z	A. X probably has an effect	B. X probably has an effect[a]
If evidence of X's effect (e.g., a significant $X{\rightarrow}Y$ association) *disappears* when controlling for Z	C. X probably has no effect	D. X may or may not have an effect; the $X{\rightarrow}Z$ confound is not disentangled

[a]The word *effect* here refers to a true causal influence rather than some empirical estimate of that influence. The phrase "X probably has an effect" is applied to both statistical and experimental controls, although the probability that statistical significance reflects an actual causal effect is greater in the context of experimental controls, assuming that random selection and random assignment mitigate confounds that might arise when X is measured rather than manipulated.

Two forms of ambiguity arise in Case D where, given the confounding, there is no way of knowing whether, or to what degree, the X-Y covariance shared with Z is due to $X{\rightarrow}Y$ causation. This problem is illustrated in the Venn diagram in Figure 3, where each circle represents the unit variance of a hypothetical variable (variance scaled to 1.0) and where overlapping regions represent shared variance (e.g., r^2). In a standard OLS run, X's effect is tested by taking the region of shared variance between X and Y that is not shared with other predictors and comparing it with error (variance in Y shared with no predictors). This tests whether X's contribution to the overall model R^2 has a unique component that itself is statistically significant. To be conservative, if this unique component is not significant even though the bivariate X-Y correlation is, the rule of thumb is to conclude that X has no effect on Y. In reality, however, this data pattern is ambiguous with respect to X's causal role because we do not know what, if any, proportion of the nonunique X-Y covariation arose from $X{\rightarrow}Y$ causation.

The second source of ambiguity takes the form of interpretive license granted to Case D. If we assume that the other criteria for mediation in Steps 1' through 3' are met (e.g., r_{xy} given W is statistically significant), we would conclude the following when adding either Z or W to the model:

1. E: If evidence of an X-Y association disappears when adding Z, X has no real effect.

2. F: If evidence of an X-Y association disappears when adding W, X has an effect, but it is mediated by W. As is widely known, but sometimes overlooked, the data cannot distinguish one model from the other, where researchers are allowed to draw diametrically opposed conclusions depending on their preferred theory. (The previous patterns are consistent with various other models as well such as that in which Z causes both X and Y.)

These two sources of ambiguity are amplified when they exist together in a single model in which one set of statistical controls is included for potential confounds (or error reduction) and another set is included for potential mediators. In Figure 2, for example, consider how our estimates of W's po-
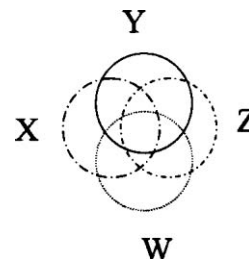

FIGURE 3   Venn diagram of unit variances among four hypothetical variables. V is the DV, X is the predictor, W is the predictor and potential mediator of $X{\rightarrow}Y$, and Z is the "other" predictor or predictors.

tential mediation change when we include Z to control for potential confounding effects:

Without Z in the Models	With Z in the Models
1. X affects Y	1. X affects Y
2. X affects W	2. X does not affect W
3. W affects Y	3. W does not affect Y
4. Most (roughly 70%) of X's effect can be accounted for by W; roughly the area $A \cap W \cap Y$ divided by the area $X \cap Y$	4. None of X's effect can be accounted for by W

Both models support X as an important predictor of Y. However, the first suggests that most of this effect is mediated by W, whereas the second suggests none of it is. The incongruity arises from the fact that including Z adjusts for potential confounding between Z and X, Z and W, and therefore Z and the covariation between X and W. In the case depicted in Figure 2, Z can account for virtually all of the shared variance between X and W and thereby eliminates the evidence of W mediating role. Conventional wisdom would suggest that we include Z and conclude that there is no evidence of a $X \rightarrow W \rightarrow Y$ path. Our intuition, however, suggests caution here because limiting our tests to mediation of X's unique component (unique from Z) ignores roughly half of X's potential effect on Y, the half that might be mediated by W. Although it is unlikely (though possible) that collinearity arising from a single additional predictor would produce such a result, the probability is much higher when we include numerous additional predictors, as has become common in marketing in which large surveys and increasingly detailed databases inflate the chances of multicollinearity (Mason & Perreault, 1991).

Finally, X may not have an effect until Z is added to the model, where Z is referred to as a *suppressor variable* (an X-Z correlation suppresses the bivariate X-Y correlation). Suppression can exist for any variable in the theorized causal path, so the validity of standard path analytic tests is threatened at every level of the path by both the inclusion and exclusion of relevant predictors (potential confounds and suppressors, respectively).

The number of possible patterns and the degree of potential ambiguity grows exponentially as we add predictors. However, researchers can simplify their assessments by not testing the effects of each additional predictor separately, but by testing them simultaneously. Given the multiple sources and general pervasiveness of ambiguity in tests of mediation, researchers are encouraged to provide a thorough report of mediation-relevant data patterns. Specifically, there are four suggestions:

1. Exercise extreme care in justifying a priori the theory and the constructs needed to test it. Given the correlational nature of the data, rarely will the data themselves discriminate between competing models, especially without experimental controls that could anchor causal direction (MacCallum,

1986). The logic of the underlying theory will then determine the study's contribution as much as will the associated data.

Researchers might justify each measure, construct, and path included in the model, indicating the role of each to (a) test a potential cause, (b) control for a potential confound, (c) uncover an effect masked by suppression, (d) reduce error variance, (e) test for mediation, and (f) test for moderation. It would also be useful to discuss things omitted from the model: (a) potentially relevant constructs and what their roles might be (confound, suppressor, error reducer, etc.) and (b) causal paths among constructs included in the model and why we should believe they do not exist. Addressing both the inclusion and exclusion of constructs and effects might require an appendix to mitigate length, but researchers might do so because this rationale will discriminate one competing model from another as much, and perhaps more so, than will any given pattern of data.

2. Publish the correlation matrix for all endogenous and exogenous indicators and constructs, preferably with covariances above the diagonal. Doing so allows readers to evaluate the raw bivariate associations uncontaminated by other variables. Researchers may also want to print path coefficients as well as bivariate correlations in their path diagrams (e.g., Law, Hawkins, & Craik, 1998).

3. Report two sets of mediation tests, one based on complete models and one based on single-path-only models (those that address only the variables involved of the focal causal path). For example, a single-path-only model that consists of X, W, and Y can be thought of as a raw trivariate association uncontaminated by other variables. The results of single-path-only and complete models will concur when neither collinearity nor multicollinearity are substantial.

4. State supporting results in an unbiased fashion. Although popular phrases such as "the data are consistent with our model" are accurate at a literal level, they risk misleading readers by implying (through omission) more empirical support than that found. The question arises as to whether the data are equally consistent with other models that may also be defendable theoretically. The bias toward confirmatory evidence is ubiquitous and difficult to irradiate even among experts (e.g., Bateman & Zeithaml, 1989). To combat this bias, researchers might use phrases that are not only linguistically accurate but that inform as well (e.g., the data are consistent with this model as well as the following models, the data are not inconsistent with our model or with others, etc.).

REFERENCES

Aiken, Leona S., & West, Stephen G. (1991). *Multiple regression: Testing and interpreting interactions.* Newbury Park, CA: Sage.

Bagozzi, Richard P. (1980). *Causal modeling in marketing.* New York: Wiley.

Baron, Reuben M., & Kenny, David A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51,* 1173–1182.

Bateman, Thomas S., & Zeithaml, Carl P. (1989). The psychological context of strategic decisions: A model and convergent experimental findings. *Strategic Management Journal, 10,* 59-74.

Blalock, Hubert M. (1971). *Causal models in the social sciences.* Chicago: Aldine.

Cohen, Jacob, & Cohen, Patricia. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences.* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Cook, Thomas D., & Campbell, Donald T. (1979). *Quasi-experimentation: Design and analysis issues for field settings.* Chicago: Rand McNally.

Davis, James A. (1985). *The logic of causal order.* Newbury Park, CA: Sage.

Heath, Timothy B. (1990). The logic of mere exposure: A reinterpretation of Anand, Holbrook, and Stephens (1988). *Journal of Consumer Research, 17,* 237-241.

Jaccard, James, Turrisi, Robert, & Wan, Choi K. (1990). *Interaction effects in multiple regression.* Newbury Park, CA: Sage.

Johnston, John. (1984). *Econometric methods.* New York: McGraw-Hill.

Judd, Charles M., & McClelland, Gary H. (1989). *Data analysis: A model comparison approach.* San Diego: Harcourt Brace Jovanovich.

Judge, George G., Griffiths, W. E., Hill, R. Carter, & Lee, Tsoung-Chao C. (1980). *The theory and practice of econometrics.* New York: Wiley.

Law, Sharmistha, Hawkins, Scott A., & Craik, Fergus I. M. (1998). Repetition-induced belief in the elderly: Rehabilitating age-related memory deficits. *Journal of Consumer Research, 25,* 91-103.

MacCallum, Robert. (1986). Specification searches in covariance structure modeling. *Psychological Bulletin, 100,* 107-120.

Maddala, G. S. (1977). *Econometrics.* New York: McGraw-Hill.

Mason, Charlotte H., & Perreault, William D. (1991). Collinearity, power, and interpretation of multiple regression analysis. *Journal of Marketing Research, 28,* 268-280.

Pedhazur, Edward J. (1982). *Multiple regression in behavioral research.* New York: Holt, Rinehart, & Winston.

Julie Irwin
University of Pennsylvania

The second part of the question raises issues regarding testing moderation (interactions). Some excellent books have been published on the subject, and we do not attempt to duplicate their thoroughness here (e.g., Aiken & West, 1991; Cohen & Cohen, 1983; Jaccard, Turrisi, & Wan, 1990; Judd & McClelland, 1989; see also Burks, 1926, 1928). However, it may be useful to present a short tutorial on the interpretation of regression coefficients when interactions are and are not included. All of these arguments hold equally well for other nonlinear equations (e.g., polynomials with higher order terms; Cohen, 1978).

The terms *moderator effect* and *interaction* are used somewhat interchangeably. The term *moderator* typically is invoked when the researcher wishes to focus on one component of an interaction. The research question of interest is whether this variable's effect on the DV is changed (exacerbated or moderated) by the level of the other component of the interaction. Statistically, though, this focus is irrelevant; interactions are symmetrical in the sense that they affect both components of the product term and may be interpreted from either perspective; if Z moderates X's effect, then so too does X moderate Z's.

Regarding the interpretation of regression coefficients, suppose one is interested in the relation between DV Y and predictor variables X and Z. The following models are of potential interest:[1] Model A, $Y = b_0 + b_1 X$. The coefficient b_1 tests if there is a relation, overall, between X and Y.

Model B, $Y = b_0 + b_1 X + b_2 Z$. The terms b_1 and b_2 are partial regression coefficients that test the effect of one variable when controlling statistically for the other. The b_1 slope will remain unchanged from Model A if there is no redundancy (collinearity)[2] between X and Z, although its statistical significance may change due to reduced error variance. Neither b_1 nor b_2 tests whether there is an interactive relation between X and Z, so it is not necessary to show that b_2 is not statistically significant or that b_1 is, to establish moderation.

Model C, $Y = b_0 + b_1 X + b_2 Z + b_3 X \times Z$. This model includes an interaction term, and the interpretation of all of the coefficients is affected, in ways that may seem counterintuitive. In fact, researchers appear to apply heuristics learned from additive regression to regression with an interactive term, often with misleading results (Irwin & McClelland, 1998). The coefficient b_1 tests the effect of X on Y when Z equals 0, and b_2 tests the effect of Z on Y when X equals 0. An algebraic reformulation of the previous equation makes clear why this is so:

$$Y = (b_0 + b_2 Z) + (b_1 + b_3 Z)X \qquad (1)$$

The b_1 test is not equivalent to the main effect test of the effect of X on Y, and to avoid confusion with the main effect test, researchers sometimes use the term *simple effect* instead of *main effect* (Aiken & West, 1991; Irwin & McClelland, 1998; analogous to a simple effect in testing contrasts in ANOVA). Note that the b_1 effect is the test when Z is zero, even if zero is not a meaningful value for Z to take (e.g., if Z was measured on a Likert scale ranging from 1 through 7). On the other hand, the variables may be linearly transformed so that zero is a useful value and the b_1 test is meaningful (e.g., Judd & McClelland, 1989; Irwin & McClelland, 1998). For instance, if the variables are mean centered (i.e., the mean is subtracted from each value), b_1 tests the effect of X on Y when Z is average. If Z is symmetrically distributed, this test will be essentially equivalent to the main effect of X on Z. By extension, contrast coded variables with a mean of zero will induce tests essentially equivalent to main effects tests in ANOVA. The main point is that in additive multiple regression lacking interaction terms, any linear recoding of variables does not change the coefficients' values or interpretations. With interaction terms, however, coding becomes important. Regardless of Z's effect

---

[1]By convention, the same coefficient is used for a given predictor across models, with the understanding that the interpretation of the coefficient can vary across models.

[2]It has recently been shown that collinearity among predictors in fact increases the power to detect interactions (McClelland & Judd, 1993).

(simple or moderating), adding an interaction term can change b_1 depending on how Z is coded and the meaning of zero on that scale. A change in b_1 (or b_2) from the simple additive model (Model B) to the model with an interaction term (Model C) has no bearing on whether there is a moderating effect of Z.

The partial regression coefficient on the interaction term is the only indication of whether there is a moderator effect. If b_3 is significant, then the effect of X on Y is different at different levels of Z. Likewise, the relation between Z and Y is different at different levels of X. These conclusions hold regardless of whether b_1 or b_2 is statistically significant.

Occasionally, researchers attempt to test for interaction effects without including the components of the interaction in the model (the main effects). For instance, they may test the following reduced model instead of the full Model C:

$$Y = b_0 + b_4 X \times Z \qquad (2)$$

This product-only model is problematic because the coefficient (b_4) on the product term is not invariant over linear transformations of X and Z. Except for very special circumstances, b_4 will not be equal to the full model's b_3.

The importance of including component terms (i.e., X and Z) when testing interactions (X^*Z) has been addressed in depth elsewhere (Cohen, 1978; Cronbach, 1987). In brief, multiplying X by Z introduces scaling artifacts that are resolved by including in the model separate effects for each of the interaction's components. For example, suppose that X and Y were coded in larger units than Z (e.g., X and Y in dollars and Z in decades). The simple relations (i.e., nonpartialed correlations) between the product term X^*Z and the variables Y, X, and Z would be affected by these scaling differences. Most important for our purposes, there would be a scale-driven relation between Y and X^*Z. Including the components of an interaction in the model removes scaling artifacts and leaves us with the test we desire: a product term that tests whether the slope of Y on X depends on the level of Z regardless of scale. Note that the b_3 test (the moderation test in the full model) is invariant over linear rescalings of X and Z, as it should be. The bottom line is that the reduced (product-only) model is not acceptable because it fails to provide a scale-independent test for changes in the X–Y and Z–Y slopes across levels of the other predictor. More generally, therefore, to ensure valid tests of higher order effects, lower order effects should be included in the model. Readers may also find interesting Henderson and Velleman (1981), LeClerc and Little (1997), Lynch (1985), and Sharma, Durand, and Oded (1981).
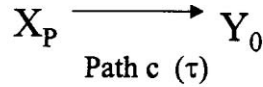
REFERENCES

Aiken, Leona S., & West, Stephen G. (1991). *Multiple regression: Testing and interpreting interactions.* Newbury Park, CA: Sage.

Burks, Barbara S. (1926). On the inadequacy of the partial and multiple correlation technique. *Journal of Educational Psychology, 17,* 532–540.

Burks, Barbara S. (1928). Statistical hazards in nature–nurture investigations. In M. Whipple (Ed.), *National society for the study of education: 27 yearbook* (Part 1, pp. 15–35). Bloomington, IL: Public School Publishing Company.

Cohen, Jacob. (1978). Partialed products are interactions; partialed powers are curve components. *Psychological Bulletin, 85,* 858–866.

Cohen, Jacob, & Cohen, Patricia. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences.* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Cronbach, Lee J. (1987). Statistical tests for moderator variables: Flaws in analyses recently proposed. *Psychological Bulletin, 102,* 418–420.

Henderson, James, & Velleman, Paul F. (1981). Building multiple regression models interactively. *Biometrics, 37,* 391–411.

Irwin, Julie R., & McClelland, Gary H. (1998). *Heuristics for moderated regression models* (Working Paper). Wharton: University of Pennsylvania Press.

Jaccard, James, Turrisi, Robert, & Wan, Choi K. (1990). *Interaction effects in multiple regression.* Newbury Park, CA: Sage.

Judd, Charles M., & McClelland, Gary H. (1989). *Data analysis: A model comparison approach.* San Diego, CA: Harcourt Brace Jovanovich.

LeClerc, France, & Little, John D. C. (1997). Can advertising copy make FSI coupons more effective? *Journal of Marketing Research, 34,* 473–484.

Lynch, John G., Jr. (1985). Uniqueness issues in the decompositional modeling of multiattribute overall evaluations: An information integration perspective. *Journal of Marketing Research, 22,* 1–19.

McClelland, Gary H., & Judd, Charles M. (1993). Statistical difficulties of detecting interactions and moderator effects. *Psychological Bulletin, 114,* 376–390.

Sharma, Subash, Durand, Richard M., & Gur-Arie, Oded. (1981). Identification and analysis of moderator variables. *Journal of Marketing Research, 28,* 291–300.

Professor Tim Ambler
London Business School

A variable is a mediator "to the extent that it accounts for the relation between the predictor and criterion" (Baron & Kenny, 1986, p. 1176). The introduction of intervening variables has proved valuable (Hulland, Chow, & Lam, 1996). For example, traditional multivariate methods showed that the size of the firm was a driver of export performance (negatively in Cooper & Kleinschmidt, 1985), though subsequent analysis gave mixed results (Naidu & Prasad, 1994). With hindsight, it seems unlikely that the firm's size can predict performance. If firms have to be big to be successful, small ones would not grow and, conversely, if they had to be small, large firms would fail. Firm size is much more likely to be a moderator.

As Baron and Kenny (1986) pointed out, intervening variables may be introduced when the relation between the IV and the DV do not accord with theoretical expectations. If the relation is unexpectedly weak, a moderator may be sought, and if it is surprisingly strong, a mediator may provide the explanation. The importance of theory in determining causality cannot be overemphasized. Virtually all the analysis that follows deals with correlation, and yet much of the literature usually states, or implies, causality. It should be determined from the context in addition to the associative tests that fol-

Model 1:

$$X_P \longrightarrow Y_0$$
Path c (τ)

Model 2:

Path a (α)↗X_M↘ Path b (β)

$$X_P \longrightarrow Y_0$$
Path c (τ')

Model 3:

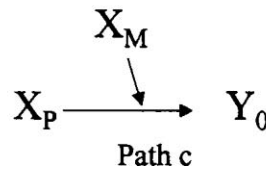$$X_M$$
$$X_P \longrightarrow Y_0$$
Path c

FIGURE 4   Models to distinguish mediators and moderators.

low. Where time series data are available, so called "Granger causality" (Ashley, Granger, & Schmalensee, 1980; Granger, 1988) infers causality where the IV occurs systematically before the DV.

A simple metaphor for the distinction between mediation and moderation uses plumbing. The original pipe allows water to flow from the IV, X_P, to the DV, Y_0, Path c in Figure 4, Model 1. If the introduction of the new variable X_M causes the water to flow along Paths a and b instead of c, as in Model 2, then X_M is a mediator. The new Paths a and b replace c. If X_M operates, however, as a valve regulating the flow from the IV to the DV, as in Model 3, then X_M is a moderator (Ambler, 1998).

Following MacKinnon, Warsi, and Dwyer (1995), let us first consider mediation. Path c in Model 1 is represented by

$$Y_0 = \tau X_P + e \qquad (1)$$

where Y_0 is the outcome variable (DV) adjusted to remove the intercept,[1] X_P is the predictor, or IV, and e is the unexplained variability.

The Baron and Kenny (1986) conditions require two regressions to test mediation:

$$X_M = \alpha X_P + e \qquad (2)$$

---

[1]This issue will become important if M is a moderator because the correlation between the independent variable (IV) and IV*M causes multicollinearity when they are used together in regression. This problem can be greatly reduced by centering the variables (i.e., subtracting the sample mean from the dependent variable; Aiken & West, 1991).

$$Y_0 = \tau' X_P + \beta X_M + e \qquad (3)$$

If the coefficients α, β, and τ are significant, and τ' equals 0, then X_M is a mediator. X_M is a partial mediator when τ is greater than τ', but τ' is still significant.

MacKinnon et al. (1995) showed that, for Equations 1 through 3
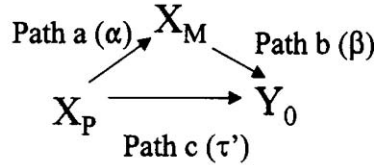
$$\tau - \tau' = \alpha\beta \qquad (4)$$

which means that we need Equations 1 or 2, but not both, because either α or β being insignificant (statistically zero) would cause τ' equals t equals 0.

The mediated effect is (τ − τ'), or αβ (MacKinnon et al., 1995). The percentage of the effect that is mediated is given by {100 × αβ/[αβ + τ']} (MacKinnon & Dwyer, 1993). Alternatively, the proportion of indirect to nonzero direct effect is given by {αβ/τ}.

Care will be needed where α, β, and τ do not all have the same sign. Suppose first that all coefficients are positive but then any one is replaced by its inverse (e.g., rainfall is replaced by drought or trust by lack of trust). Two paths change from positive to negative correlations. Each time a single variable reverses polarity, two relations reverse polarity—that is, αβτ remains positive. As τ' is likely to become insignificant, its sign is irrelevant. Conversely, suppose that one path was negative (e.g., the original Path c before the intervention of the mediator, i.e., τ). Then, a negative relation is being mediated (replaced) by two positive ones, which is impossible. Accordingly, αβτ is greater than 0 is a condition for mediation that has not previously appeared in the literature, as far as I know.

Sobel (1982) derived the usual approximate significance test for the mediated effect (αβ) giving the SE as $\sigma_{\alpha\beta} = \sqrt{\sigma_\alpha^2 \beta^2 + \sigma_\beta^2 \alpha^2}$, omitting the $\sigma_\alpha^2 \sigma_\beta^2$ term as being small. Sample sizes need to be above 50 for reasonable and around 200 for accurate mediated effect SE estimates (MacKinnon & Dwyer, 1993).

Let us turn to the issue of moderation. "A moderator is a qualitative (e.g., sex, race, class) or quantitative (e.g., level of reward) variable that affects the direction and/or strength of the relation between an independent or predictor variable and a dependent or criterion variable" (Baron & Kenny, 1986, p. 1174). To test for moderation, we maintain Equations 1 and 2 (presented earlier) and replace Equation 3 with

$$Y_0 = \tau' X_P + \beta X_M + \gamma(X_P * X_M) + e \qquad (5)$$

If α equals β equals 0 (neither are significant) but γ is significant, then X_M is a pure moderator (Sharma, Durand, & Gur-Arie, 1981; i.e., it does not covary with either the IV or DV).

Sharma et al. (1981) noted confusion in the literature over whether moderators can be related to IVs or the DV and sought to deal with the problem by allowing two weaker

forms of moderation: quasi moderation and homologizers. Quasi moderation relaxes the requirement that X_M is related to neither the DV nor IV. Thus, in Equations 2 and 5, α, β, or both may be significant. If β is significant, Equation 5 would become symmetric between the moderator and the IV. The resulting ambiguity over which was the IV and which was the moderator led Sharma et al. to conclude (p. 294) that even a quasi moderator must not be related to the DV (i.e., $\beta = 0$). A homologizer variable is not related to the IV or DV either, but unlike a pure moderator, has no interaction with the IV (i.e., γ in Equation 5 is not significant).

In conclusion, the tests for mediation and moderation are set out in Table 1. These indicate that the well-known Baron and Kenny (1986) tests for mediation are overspecified in one respect (a does not need to be tested) and underspecified ($\alpha\beta\tau > 0$) in another. Their nonuse of equations also leads to some ambiguity which this article set out to resolve.

TABLE 1
Statistical Tests for Intervening Variables

Variable	Condition 1: Improved Fit	Condition 2: β and γ Significant?	Condition 3: τ Significant?	Condition 4: $\alpha\beta\tau > 0$
Mediator	✓	β but not γ	$\tau > \tau'$, perfect mediation if τ' not significant	✓
Moderator	✓	γ but not β	$\tau \neq \tau'$	not an issue

REFERENCES

Aiken, Leona S., & West, Stephen G. (1991). *Multiple regression: Testing and interpreting interactions.* Newbury Park, CA: Sage.

Ambler, Tim. (1998). *Mediation and moderation: Roles and tests* (Working Paper No. 498–903). London: London Business School.

Ashley, R., Granger, Clive W. J., & Schmalensee, Richard. (1980). Advertising and aggregate consumption: An analysis of causality. *Econometrica, 48,* 1149–1167.

Baron, Reuben M., & Kenny, David A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51,* 1173–1182.

Cooper, Robert G., & Kleinschmidt, Elko J. (1985). The impact of export strategy on export sales performance. *Journal of International Business Studies, 3,* 7–55.

Granger, Clive W. J. (1988). Some recent developments in a concept of causality. *Journal of Econometrics, 39,* 199–211.

Hulland, John, Chow, Yiu Ho, & Lam, Shunyin. (1996). Use of causal models in marketing research: A review. *International Journal of Research in Marketing, 13,* 181–197.

MacKinnon, David P., & Dwyer, James H. (1993). Estimating mediated effects in prevention studies. *Evaluation Review, 17,* 144–158.

MacKinnon, David P., Warsi, Ghulam, & Dwyer, James H. (1995). A simulation study of mediated effect measures. *Multivariate Behavioral Research, 30,* 41–62.

Naidu, G. M., & Prasad, V. Kanti. (1994). Predictors of export strategy and performance of small- and medium-sized firms. *Journal of Business Research, 31,* 107–115.

Sharma, Subhash, Durand, Richard M., & Gur-Arie, Oded. (1981). Identification and analysis of moderator variables. *Journal of Marketing Research, 18,* 291–300.

Sobel, Michael E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. In Samuel Leinhardt (Ed.), *Sociological methodology* (pp. 290–293). San Francisco: Jossey-Bass.