Mediators, Moderators, and Tests for Mediation

Lawrence R. James
Georgia Institute of Technology

Jeanne M. Brett
Graduate School of Management,
Northwestern University

The following points are developed. First, mediation relations are generally thought of in causal terms. Influences of an antecedent are transmitted to a consequence through an intervening mediator. Second, mediation relations may assume a number of functional forms, including nonadditive, nonlinear, and nonrecursive forms. Special attention is given to nonadditive forms, or moderated mediation, where it is shown that, although mediation and moderation are distinguishable processes, a particular variable may be both a mediator and a moderator within a single set of functional relations. Third, current procedures for testing mediation relations in industrial and organizational psychology need to be updated because these procedures often involve a dubious interplay between exploratory (correlational) statistical tests and causal inference. It is suggested that no middle ground exists between exploratory and confirmatory (causal) analysis and that attempts to explain how mediation processes occur require well-specified causal models. Given such models, confirmatory analytic techniques furnish the more informative tests of mediation.

Researchers in industrial and organizational psychology and organizational behavior are placing increasing emphasis on studying mediation models in which the influence of an antecedent is transmitted to a consequence through an intervening mediator. Cases in point include (a) job-perception studies in which the effects of work environments (antecedents) are transmitted to affective and behavioral outcomes (consequences) by intervening job perceptions (mediators; cf. Brass, 1981; Oldham & Hackman, 1981; Rousseau, 1978a, 1978b; Sutton & Rousseau, 1979); (b) attrition studies in which the effects of environmental events and individual attributes are transmitted to attrition behaviors via intervening behavioral intentions to stay or leave (cf. Arnold & Feldman, 1982; Hom & Hulin, 1981; Horn, Katerberg, & Hulin, 1979; Miller, Katerberg, & Hulin, 1979; Mobley, Hand, Baker, & Meglino, 1979; Mobley, Horner, & Hollingsworth, 1978); and (c) attribution research in leadership, where the effects of subordinate performance on subsequent behaviors by the leader toward a subordinate are transmitted by the leader's attributions of the causes of the subordinate's performance (cf. Ilgen, Mitchell, & Fredrickson, 1981; McFillen, 1978; Mitchell & Kalb, 1981; Mitchell & Wood, 1980).

At the theoretical level, studies such as these are typically based on causal models that assume complete mediation as well as additive and linear causal relations. To illustrate the principles involved, a complete mediation model has the form \( x \rightarrow m \rightarrow y \), where \( x \) is the antecedent, \( m \) is the mediator, and \( y \) is the consequence. The antecedent \( x \) is expected to affect the consequence \( y \) only indirectly through transmission of influence from \( x \) to \( y \) by the mediator \( m \). The indirect transmission of influence from \( x \) to \( y \) via \( m \) denotes that all of the effect of \( x \) on \( y \) is transmitted by \( m \). In causal terminology, this state of affairs is

Support for this project was provided under Office of Naval Research Contracts N00014-80-C-0315 and N00014-83-K-0480, Office of Naval Research Projects NR170-904 and NR475-026. Opinions expressed are those of the authors and are not to be construed as necessarily reflecting the official view or endorsement of the Department of the Navy.

The authors wish to thank Robert G. Demaree, Jeanne L. Dugas, Sigurd B. Gustafson, Allan P. Jones, Stanley A. Mulaik, and Gerrit Wolf for their helpful suggestions and advice.

Requests for reprints should be sent to Lawrence R. James, School of Psychology, Georgia Institute of Technology, Atlanta, Georgia 30332.
directly to \( m \) is held constant, if these when \( y \) —» which is: subordinate performance \( (x) \)

Assuming linear and additive causal relations, the complete mediation model thus predicts that \( x \) has a direct effect on \( m \), \( m \) has a direct effect on \( y \), and \( x \) is not related directly to \( y \) when \( m \) is held constant. If these predictions are empirically confirmed, then one may infer that the complete mediation model has been corroborated and therefore is useful for attempting to explain how \( x \) is related to \( y \) through the intervening mediator \( m \) (James, Mulaik, & Brett, 1982). Explanation is a matter of elucidating the processes by which \( m \) is a linear, additive function of \( x \) and \( y \) is a linear, additive function of \( m \) (Rozeboom, 1956).

This article has two objectives, both of which resulted from observations that many studies that ostensibly derived from a linear, additive, complete mediation model departed from this theoretical base during empirical operationalizations of the model and/or explanations of the results of empirical tests of the model. The first objective is to discuss complete mediation models that imply additivity but take on a distinctly nonadditive flavor in empirical operationalizations, empirical tests, and explanations of results. A case in point is the attribution model of leader behavior proposed by Green and Mitchell (1979), which is: subordinate performance \( (x) \) → leader’s causes of the attributions of the causes of the subordinate’s performance \( (m) \) → leader behavior toward the subordinate \( (y) \). This model appears to assume the \( x \) → \( m \) → \( y \) form, where attributions transmit, additively and linearly, influence from subordinate performance to leader behaviors. Some attribution studies in leadership do indeed maintain an additive, linear, complete mediation form, although causes of leaders’ attributions other than subordinate performance are typically included in investigations (e.g., interdependence of supervisor and subordinate, see prior references). In other operationalizations, tests, and interpretations of the complete mediation model, the attributions are not treated as simple mediators. Rather, they appear to assume the role of moderators (Ilgen & Knowlton, 1980; Knowlton & Mitchell, 1980—see also Goodstadt & Kipnis, 1970; Kipnis & Cosentino, 1969; Kipnis, Silverman, & Copeland, 1973). For example, empirical evidence indicates that if poor subordinate performance is attributed to lack of effort, but not to lack of ability, then the leader is likely to increase close supervision and decrease support. Conversely, if poor subordinate performance is attributed to lack of ability (interpreted here as lack of training and experience), but not to lack of effort, then the leader is likely to increase support but not close supervision, at least not in the sense of the use of coercive power (e.g., reprimand the subordinate).

These informative findings imply that the comparative strengths of attributions to ability and effort serve to moderate the relation between subordinate performance and leader behavior. This stimulates the question: Are the attributions also mediators in the sense that they intervene between subordinate performance and leader behavior toward the subordinate, as predicted by the Green and Mitchell (1979) model? An answer to this question is not easily furnished because it requires that we explore relations between the concepts of mediator and moderator. In the broader context, of concern are answers to questions such as, “Must mediation relations be additive?” “May mediators also be moderators?” and “May moderators also assume the role of mediators?” Exploration of the mediator and moderator concepts and answers to these and related questions comprise the first objective of this article.

The second objective is to demonstrate why investigators should devote more attention to the assumptions for confirmatory (causal) analysis before conducting confirmatory tests of complete mediation models. Consider, for example, the job perception and attrition studies cited at the beginning of this article. Each of these studies proposed a verbal, and often graphic, complete mediation model, which was then tested using analytic procedures typically associated with exploratory (i.e., correlational) analysis, such as hierarchical regression and/or partial correlation. When methods such as hierarchical regression and partial correlation are used to test complete mediation hypotheses, it follows that the methods have assumed the roles of confirmatory tests. It follows also that, like other forms of confirmatory analysis (e.g., path analysis), these methods should only be employed after conditions for confirmatory analysis have been reasonably satisfied. The use of
traditionally exploratory methods to test causal models does not absolve the researcher from having to satisfy conditions for confirmatory analysis.

To set a fair stage for discussion, it should be noted that prior tests of complete mediation models in the job perception and attrition literatures represent initial attempts to advance from purely exploratory forms of analysis to confirmatory tests of causal hypotheses. Furthermore, investigators typically devoted attention to some of the conditions for confirmatory analysis, such as justifying the presumed causal ordering among variables. Our concern is that these initial and much needed attempts to advance from exploratory analysis to confirmatory analysis must now be regarded as incomplete in the context of recent accumulation of knowledge in industrial and organizational psychology regarding all of the conditions that are prerequisite to meaningful confirmatory analysis (James et al., 1982). Moreover, whereas hierarchical regression and partial correlation may indeed be used in the confirmatory mode (Cohen & Cohen, 1983), these methods are limited in regard to both the types of causal models for which they are applicable and the information they provide (Griffin, 1977).

To illustrate concerns pertaining to conditions for confirmatory analysis, consider that empirical support for the complete mediation model, job environment (e.g., job technology) \(\rightarrow\) job perceptions (e.g., job challenge) \(\rightarrow\) job satisfaction, is interpreted to mean that (a) job perceptions transmit causal influences from the job environment to job satisfaction, and (b) individuals experiencing the same or similar type(s) of environments may differ in terms of how they perceive the job and, therefore, differ in how they respond affectively to the job (see prior references). Interpretation “a” denotes that job perceptions covary significantly with \(between-job\) variation in such things as levels of technology and that covariation in job perceptions is associated significantly with covariation in job satisfaction. The empirical data support this interpretation, which reflects an attempt to enhance explanation of the processes by which job environments influence job satisfaction by identifying an intervening, perceptual mediator(s).

Interpretation “b” is not a legitimate causal inference because relevant causes of reliable \(within-job\) variation in job perceptions (and job satisfaction) are not included in the causal models or tested empirically. Clearly, if job perceptions and job satisfaction vary reliably within levels of technology, then the intervening job perceptions are not just transmitting influences from job technology to job satisfaction. Rather, other causes of job perceptions (and job satisfaction), such as personal attributes and social influences, will likely have to be invoked to explain the reliable \(within-job\) variation in job perceptions (see James & Jones, 1980; Kim, 1980; O'Reilly, Parlette, & Bloom, 1980; Schmitt, Coyle, White, & Rauschenberger, 1978; Thomas & Griffin, 1983). When not included in a causal model, these other causes are referred to as \(unmeasured variables\). Given stipulations to be discussed later, failure to include one or more unmeasured variables in the causal model results in biased statistical results and erroneous causal inferences in regard to relations among variables included explicitly in the causal model (cf. James, 1980).

Unmeasured variable problems are symptomatic of the incomplete transition from exploratory modes of analysis to confirmatory modes of analysis. In the presentation of Objective 2, we will address these problems and other key conditions required for confirmatory analysis that must be considered in order to effect a complete transition from exploratory analysis to confirmatory analysis. We will also recommend that hierarchical regression and partial correlation should not be used in place of confirmatory analytic procedures.

Objective 1: An Attempt to Distinguish Between Mediators and Moderators

The first objective of this article is to define mediation and moderation and then to compare mediation with moderation. As part of this process, we shall see that contemporary definitions of mediation are somewhat misleading and that the distinction between mediation and moderation can be blurred at both the theoretical and operational levels of explanation.

Contemporary Definitions

The definition of mediator advanced by Rozeboom (1956) for hypothetical constructs appears to be characteristic of the linear, ad-
ditive, complete mediation models employed in many areas of psychology and the social sciences. This definition is as follows: \( m \) is a mediator of the probabilistic relation \( y = f(x) \) if \( m \) is a probabilistic function of \( x \) (i.e., \( m = f(x) \)) and \( y \) is a probabilistic function of \( m \) (i.e., \( y = f(m) \)), where \( x \), \( m \), and \( y \) have different ontological content (i.e., represent different hypothetical constructs or latent variables). As discussed earlier, theoretical operationalizations of mediation are usually based on causal mediation models, which in shorthand notation assume the form \( x \rightarrow m \rightarrow y \). In addition to the obvious point that a causal order must be assumed, the typical causal mediation model is based on the premises that (a) the \( fs \) in \( m = f(x) \) and \( y = f(m) \) represent linear, additive, and recursive (i.e., unidirectional) functions, which in equation form for deviation scores are \( m = bx + e \) and \( y = bm + e \), where \( b \) is a causal parameter and \( e \) is an error or disturbance; (b) \( m \) transmits all of the influence of an antecedent \( x \) to a consequence \( y \), which implies that \( x \) and \( y \) are indirectly related and that the relation between \( x \) and \( y \) vanishes if \( m \) is held constant; and (c) the inclusion of \( m \) in the model serves to enhance the explanatory power of the model because \( m \) furnishes substantive explication of how the antecedent is related to the consequence, whereby "related" means how \( x \) "produces," "acts on," or otherwise influences \( y \) (cf Alwin & Hauser, 1975; Blalock, 1982; Cook & Campbell, 1979; Duncan, 1975; Heise, 1975; James et al., 1982; Kenny, 1979).

With respect to moderation, a variable \( z \) is a moderator if the relationship between two (or more) other variables, say \( x \) and \( y \), is a function of the level of \( z \). This definition indicates an \( x \) by \( z \) interaction, or a nonadditive relation, where \( y \) is regarded as a probabilistic function of \( x \) and \( z \). Specifically, the probabilistic function is \( y = f(x, z) \), the function \( f \) being \( y = bx + bz + b_{xz}z + e \) for deviation scores and a model linear in the parameters. The \( bs \) in this function will be regarded as noncausal, statistical parameters for the present.

If we compare this definition to the Rozeboom (1956) definition for mediation, it would seem that a number of clear lines of demarcation exist between the terms mediator and moderator. In particular, the moderator model is represented by a single, nonadditive, linear function (although often tested by a hierarchical process) in which it is desirable to have minimal covariation between the moderator and both the independent and dependent variables (Abrahams & Alf, 1972). In comparison, mediation models must be represented by at least two additive, linear functions in which it is desirable to have high degrees of covariation between the mediator and both the antecedent(s) and consequence(s). Use of the terms independent and dependent in moderator models, and antecedent and consequence in mediator models, is purposeful and indicates that moderation carries with it no connotation of causality, although a causal relation may be moderated (cf. Stolzenberg, 1979). Mediation, on the other hand, implies at the minimum a causal order, and often additional causal implications are required to explain how mediation occurred (these implications are considered later). Because of the causal overtones in mediation relations, a confirmatory analytic approach is employed below to illustrate additional issues in moderation and mediation, although the basic statistical arguments generalize to exploratory designs.

**Moderated Mediation**

Things are not necessarily as straightforward as the above definitional demarcations suggest, one reason being that mediation relations may involve a moderator, in which case the mediation relations cannot be additive. The issues here will be presented by way of illustration for a self-attribution model based on simplified and overdramatized abstractions from Bandura (1977, 1978), Jones (1973), and Weiner (1979). Suppose we conduct a study designed to test the propositions that (a) effort attributions mediate the relation between level of poor performance and degree of intended persistence for high-self-esteem individuals and (b) ability attributions mediate the relation between level of poor performance and degree of intended persistence for low-self-esteem individuals. The proposed causal models are shown in Figure 1, section a. Individuals are first given a self-esteem questionnaire and then blocked (subgrouped) into high self-estees or low self-estees, the criterion for blocking being whether an individual scores above or below a theoretical point on the self-esteem scale. Second, within the high- and low-self-
esteem blocks, individuals are assigned randomly to five bogus performance-feedback conditions (explained below). Third, individuals in all five conditions are asked to perform the same moderately difficult task, which requires mental effort and approximately 15 minutes to complete. Fourth, following task completion, individuals are given bogus performance feedback implying that they have failed the task. Degree of failure is varied on an approximately interval scale (e.g., Condition 1: 50% of the people did better than you. Condition 5: 90% of the people did better than you). Fifth, individuals are asked to make two attributions, one regarding the degree to which their performance was due to lack of effort and one regarding the degree to which their performance was due to lack of ability (e.g., 0 = Effort [ability] had no effect on my performance. . . 6 = My performance was strongly affected by a lack of effort [ability]). After completing the attribution items, individuals are asked to report the extent to which they would now be willing to participate in a similar task (e.g., 1 = Definitely not participate . . . 3 = Ambivalent about participation. . . 5 = Definitely participate). Scores on this scale represent degrees of “intended persistence.” This is checked empirically by conducting a second task, but we shall use the intended persistence indicator in order to stay in the parametric realm and thereby not get bogged down in extraneous statistical issues. Finally, the experiment is ended by debriefing participants.

To demonstrate principles, realizing that dichotomous blocking on a continuous self-

\[ \text{Figure 1} \quad \text{Mediation relations for performance feedback (PF), effort (E) and ability (A) attributions, and intended persistence (IP), moderated by self-esteem (HSE = high self-esteem, LSE = low self-esteem)} \]
estee variable is questionable and that the
relations to be presented are overdramatic, let
us suppose that the results of our study cor-
respond to a priori predictions and are as
shown in Figure 1, sections b through e. These
figures portray regression slopes associated
with correlations among raw or deviation scores
on the variables. For high self-esteem individ-
uals, Figure 1, sections b and c, suggests a
tendency to attribute increasing degrees of
failure to a steadily increasing lack of effort,
but not to ability. Specifically, scores on effort
attributions vary from 2 to 6 and are associated
with performance feedback (see Figure 1, sec-
tion b), whereas scores on ability attributions
vary randomly between 0 and 1 and are not
associated with performance feedback (see
Figure 1, section c). The explanation for these
results is that high-self-esteem individuals have
confidence in their abilities and thus are prone
to attribute unexpected failure to an unstable
cause such as lack of effort. Continuing with
high-self-esteem individuals, Figure 1, section
d, indicates that the higher the perceived lack
of effort, the more likely the intended persis-
tence to participate on a second task. The ra-
tionale here is that comparatively stronger ef-
fort attributions reflect a greater imbalance
between a positive self-concept and perfor-
mance feedback, and therefore a stronger force
to correct the imbalance by performing suc-
cessfully on the second task. Finally, inasmuch
as performance was essentially not attributed
to ability, ability attributions are unrelated to
intended persistence for high-self-esteem in-
dividuals (see Figure 1, section e).

In regard to low-self-esteem individuals,
Figure 1, sections b and c, shows that ability,
but not effort, attributions are a positive func-
tion of performance feedback. That is, the
higher the failure, the stronger the attribution
to lack of ability, for which scores vary from
2 to 6 (see Figure 1, section c), but scores on
effort attributions assume random values be-
tween 0 and 1 and are unrelated to degree of
failure (see Figure 1, section b). Figure 1, sec-
tion e, demonstrates that intended persistence
is an inverse function of ability attributions
(i.e., the stronger the attribution to lack of
ability, the lower the intention to participate
in the second task). Effort attributions are not
related to intended persistence (see Figure 1,
section d) because performance feedback was
essentially not attributed to effort. The ratio-
nale for these relations is that (a) implied fail-
ure is consistent with low-self-esteem individ-
uals' lack of self-confidence, thereby resulting
in the performance feedback — ability attri-
bution relation and (b) intent to persist, which
is never high, decreases as attributions to lack
of ability increase because individuals perceive
an increasing likelihood of failure and, as a
form of defense, withdraw to protect an al-

Now let us play the game of find the mod-
erator(s) and the mediator(s). Application of
the Rozeboom (1956) definition for mediation
indicates that self-esteem is not a mediator
because self-esteem is not a direct or indirect
function of performance feedback. That is, if
x = performance feedback and m = self-es-
teeem, then self-esteem fails to satisfy the first
criterion for mediation because m \neq f(x). This
is clearly the case because self-esteem was
measured before the experiment. Rather, self-
esteeem is a moderator, which is evident in
Figure 1 because relations between the ante-
cedent performance-feedback conditions and
the attributions, and between the attributions
and intended persistence, are contingent on
the level of self-esteem.

In contrast, the attributions appear to be
mediators. The attributions have ontological
content that differs from performance feed-
back and intended persistence (and self-es-
teeem). There is an explicit causal order in
which the attributions occur after performance
feedback and prior to intended persistence,
and, contingent on the level of self-esteem, the
attributions are effects of performance feed-
back and causes of intended persistence. This
suggests that inclusion of the attributions in
the model helps to explain how performance
feedback influences intended persistence in the
sense that the attributions specify the processes
by which the influences of performance feed-
back are transmitted to intended persistence.
Finally, the attributions are complete media-
tors of the relation between performance feed-
back and intended persistence, contingent on
the level of self-esteem, which is to say that
performance feedback affects intended persis-
tence only indirectly through the attributions.
The fact that the mediation relations are
contingent on the level of self-esteem suggests
the need to amend Rozeboom's (1956) defi-
nition of mediation to include moderation. This is easily accomplished by mapping the Rozeboom (1956) functional relations into the relations and accompanying functional equations implied by Figure 1, only here we will include the nonadditive relations required by the self-esteem moderator. The term functional equation refers to a quantitative statement of the presumed structure of causal relations among a set of variables in a self-contained system, whereby self-contained is meant that all relevant causes of an effect or endogenous variable are included in the equation for that variable (James et al., 1982; Simon, 1952, 1953, 1977). For the Rozeboom function \( \hat{m} = f(x) \), we have \( \hat{E} = f(PF, SE) \) and \( \hat{A} = f(PF, SE) \), where \( E \) = effort attribution, \( PF \) = performance feedback, \( SE \) = self-esteem, and \( A \) = ability attribution. The "f" in the functions for \( E \) and \( A \) represents a nonadditive, although linear, relation, as seen by the inclusion of interaction terms in the following functional equations for \( E \) and \( A \) (the variables in these equations and all remaining equations are assumed to be in deviation form).

\[
E = b_{E,PF}PF + b_{E,SE}SE
+ b_{E,(PF \times SE)}(PF \times SE) + e \quad (1)
\]

\[
A = b_{A,PF}PF + b_{A,SE}SE
+ b_{A,(PF \times SE)}(PF \times SE) + e \quad (2)
\]

The "be" in Equations 1 and 2 represent causal or structural parameters. For example, \( b_{E,PF} \) in Equation 1 is defined as the unique amount of change in \( E \) brought about by a unit of change in \( PF \). Given reasonable satisfaction of the assumptions or conditions for confirmatory analysis, which are discussed in Objective 2, the structural parameters may be estimated by unstandardized, ordinary least squares (OLS) regression weights. In this sense, the statistical estimating equations for Equations 1 and 2 may be thought of as simple multiple regression equations. Inspection of Figure 1, Sections b and c, indicates that estimates of the structural parameters representing the interactions (i.e., \( b_{E,(PF \times SE)} \) and \( b_{A,(PF \times SE)} \)) will be significant. In other words, the relation between \( E \) and \( PF \) is moderated by \( SE \), as is the relation between \( A \) and \( PF \). (A point worthy of brief mention is that errors of estimate as well as slope coefficients vary as a function of \( SE \) blocks in Figure 1, Section b through Section e. Technically, heterogeneous errors of estimate would preclude tests of slope—Gulliksen & Wilks, 1950).

The salient point here is that moderation may be functionally involved in the first-stage of a mediation relation, but the moderator is not a mediator. Specifically, variation in performance feedback affects only an attribution, but the explanation of the effects of performance feedback on ability and effort attributions is contingent on the level of self-esteem. Moderation carries over into the second stage of mediation in this model, because the relations between intended persistence and both effort and ability attributions are contingent on the level of self-esteem (see Figure 1, Sections d and e). Rozeboom's second functional relation for mediated relations, \( \hat{y} = f(m) \), stated separately for \( A \) and \( E \), is \( \hat{IP} = f(A, SE) \) and \( \hat{IP} = f(E, SE) \), where \( IP = \) intended persistence, and the functions again represent nonadditive relations. In equation form, the functions are as follows:

\[
IP = b_{IP,E}E + b_{IP,SE}SE
+ b_{IP,(E \times SE)}(E \times SE) + e \quad (3)
\]

\[
IP = b_{IP,A}A + b_{IP,SE}SE
+ b_{IP,(A \times SE)}(A \times SE) + e \quad (4)
\]

Like Equations 1 and 2, OLS estimates of the structural parameters representing the interactions would be significant. This suggests that the attributions transmit the influence of performance feedback to intended persistence and enhance the explanatory power of the model by specifying the processes through which feedback acts on intentions. However, such transmission and enhancement is contingent on the level of self-esteem, and although self-esteem transmits nothing from feedback to intentions, and thus cannot be a mediator, it contributes directly to the explanatory power of the model. It might also be noted that a single equation for \( IP \) could be developed. Analyses would demonstrate that the equation with the best fit to the data would involve the first-order interactions \( b_{IP,(E \times SE)}(E \times SE) \) and \( b_{IP,(A \times SE)}(A \times SE) \). Interactions involving \( (A \times E) \) and \( (A \times E \times SE) \) would be redundant with the first-order interactions using \( SE \) as the moderator.
A final test of the model would consist of ascertaining whether all of the influence of performance feedback (PF) on IP is transmitted by the mediating attribution variables. The many options available for this test, typically referred to as a "goodness of fit test" or a "test of logical consistency," include an omitted parameter test (Duncan, 1975; James et al., 1982; Namboodiri, Carter, & Blalock, 1975), a disturbance term regression test (James & Jones, 1980), and hierarchical OLS in the confirmatory mode. Given high correlations between PF and E in the high-self-esteem block, and between PF and A in the low-self-esteem block, use of the omitted parameter test would likely be subject to multicollinearity. Thus, the latter two procedures would be the prime candidates for the goodness-of-fit test. To illustrate the use of hierarchical OLS in the confirmatory mode, the regressions indicated by Equations 3 and 4 would be conducted and $R^2$'s estimated. These are referred to as $R^2_3$ and $R^2_4$ to indicate estimates based on Equations 3 and 4, respectively. Next, PF and (PF X SE) would be added to Equation 3 as independent variables, and a new $R^2$ computed, which is designated $R^2_{3+}$. A nonsignificant difference between $R^2_3$ and $R^2_{3+}$ would imply that, within the self-esteem blocks, PF is not directly related to IP when E is held constant. The key inference would be that E completely mediates the effects of PF on IP for high-self-esteem individuals. A similar process would be conducted for Equation 4, namely PF and (PF X SE) would be added to Equation 4 and $R^2_{4+}$ computed. A nonsignificant difference between $R^2_4$ and $R^2_{4+}$ would confirm the prediction that A completely mediates the influences of PF on IP for low-self-esteem individuals. Should $R^2_{3+} > R^2_3$, and/or $R^2_{4+} > R^2_4$, then at least one of the predictions based on the causal model has been disconfirmed. The resulting inference would be that at least one of the mediators is not a complete mediator, which is to say that PF has a direct effect on IP in the high-self-esteem block and/or the low-self-esteem block. In sum, moderators and mediators have different roles, even though they may occur jointly in the same model. If one is willing to adopt the formal definition of mediator for hypothetical constructs advanced by Rozeboom (1956), then specific criteria must be satisfied before a variable may be designated a mediator. It is particularly important to recognize that $m = f(x)$ and $y = f(m)$ not only assume an explicit causal order but also imply active causal processes in which m transmits the effects of x to y and, as part of this transmission process, enhances explanation because it specifies the processes by which x acts on or produces y. On the other hand, there is no requirement that mediation relations be additive. Nonadditive relations require the addition of a moderator for either the $m = f(x)$ or $y = f(m)$ relations, or both (as shown here). In this condition, the moderator is added to the function (e.g., $IP = f(A, SE)$) and "f" is specified as nonadditive. The term moderated mediation is suggested for such models to denote that mediation relations are contingent on the level of a moderator.

Roles of Variables in Mediation and Moderation

It follows from the discussion above that mediators are distinguished from moderators by the operational roles played by variables in functional relations and equations. A seemingly logical deduction is that a particular variable can be unambiguously classified as either a mediator or a moderator. In some, and perhaps most, cases this is true. In other cases it is false because a particular variable may assume the roles of both mediator and moderator in the same model, and even in the same functional relation and equation. To see how this could occur, suppose we conduct the same experiment as described above, only this time we randomly assign individuals to the five performance-feedback conditions without measurement or blocking on self-esteem. Assuming that high-self-esteem individuals are as likely as low-self-esteem individuals to be randomly placed in each performance-feedback condition, we would find that each attribution variable serves as both a mediator and a moderator.

Illustrations of the relations are presented in Figure 2. Figure 2, sections b and c, shows that ability attributions moderate the regressions of effort attributions on performance feedback and intended persistence on effort attributions. The rationale here is the same as that for self-esteem, only here high-self-esteem
individuals are represented by scores of 0 or 1 on the ability attribution scale, and low-self-esteem individuals are represented by scores equal to or greater than 2 on the ability attribution scale. The salient points are that effort attributions are mediators, whereas ability attributions are moderators. Consistent with these points is the observation that an attempt to fit a linear, additive, mediation model to the relations involving effort attributions, namely $E = f(PF)$ and $IP = f(E)$, would fail because the errors of estimate are heteroscedastic in both relations. This is easily seen, for example, in Figure 2, section a, where the cluster of points in the bivariate scatterplot would be roughly triangular if moderation by ability was disregarded.

We can now reverse the process, so to speak, and regard effort attributions as the moderator and ability attributions as the mediator. As seen in Figure 2, section b, the regression of ability attributions on performance feedback is "significant" for individuals with scores of 0 and 1 on the effort attribution scale (low-self-esteem individuals), and "nonsignificant" for individuals with scores equal to or greater than 2 on the effort attribution scale (high-self-esteem individuals). The moderation by effort attributions carries over to the regression of intended persistence on ability attributions (see Figure 1, section d).

Algebraic expression may help to clarify the points above. Mapping the Rozeboom (1956) relation $\hat{y} = f(x)$, amended for moderation, into the relations above furnishes the following functional equations:

\[
E = b_{E,PF}PF + b_{E,A}A \\
\quad + b_{E,(PF \times A)}(PF \times A) + e \\
A = b_{A,PF}PF + b_{A,E}E \\
\quad + b_{A,(PF \times E)}(PF \times E) + e
\]

Figure 2, sections a and b, denotes that OLS estimates of terms representing the interactions will be significant, thus indicating that both $A$ and $E$ assume the functional role of moderator in one of the equations. Yet, each attribution satisfies the first criterion for moderated mediation in the equation in which it serves as an endogenous (dependent) variable.

Joint roles as a mediator and as a moderator are even more apparent when Rozeboom's second criterion for a mediation relation, $\hat{y} = f(m)$ amended for moderation, is mapped into
our example. The equation is the same for either $A$ or $E$ as a mediator and/or moderator, and is as follows:

$$IP = b_{IP,E}E + b_{IP,A}A$$

$$+ b_{IP(EXA)}(E \times A) + e \quad (7)$$

Given that the estimate of $b_{IP(EXA)}$ is significant, Equation 7 may be interpreted as (a) the effects of $E$ on $IP$ are contingent on the value of $A$ (see Figure 2, section c), or as (b) the effects of $A$ on $IP$ are contingent on the level of $E$ (see Figure 2, section d). Combining the first interpretation of Equation 7 with Equation 5 gives us $E$ as a mediator and $A$ as a moderator. Combining the second interpretation of Equation 7 with Equation 6 gives us $A$ as a mediator and $E$ as a moderator.

In conclusion, it may be impossible to classify a particular variable as either a mediator or a moderator because this variable may play both roles in a set of simultaneous equations designed to represent a causal model or system (i.e., Equations 5, 6, and 7 represent a set of simultaneous, functional equations for one causal system—cf. Simon, 1977). This need not be confusing if one remembers that it is the role or roles that a variable plays that determine whether it is a moderator, a mediator, or both. Thus, applying the definitions for mediation, moderation, and moderated mediation to the operational role(s) played by a variable in each functional relation and equation over the set of relations and equations in a causal system furnishes the basis for ascertaining whether the variable is a mediator, a moderator, or both a mediator and a moderator.

**Other Amendments to the Functional Definition of Mediation**

In addition to moderation, it is necessary to extend the Rozeboom (1956) functional definition of mediation to other types of functional relations. First, there is the question of nonlinearity in the variables. For example, $m$ may be a linear, additive function of $x$, but $y$ may be an additive, nonlinear function of $m$. The mediation relation takes a form such as $\hat{m} = f(x)$, $\hat{y} = f(m^2)$, which may be tested empirically by applying hierarchical OLS procedures to operationalized functional equations (Stolzenberg, 1979). Second, mediation functions may involve nonrecursive relations, such as $x \rightarrow m \leftrightarrow y$, where "\leftrightarrow" denotes reciprocal causation. The mediation relation in this case would be $\hat{m} = f(x, y)$, $\hat{y} = f(m)$, although additional exogenous causes of $m$ and $y$ would have to be added before empirical tests are possible (cf. James & Singh, 1978). Examples of tests for mediation in nonrecursive designs are presented in James and Jones (1980) and Maruyama and McGarvey (1980). Cyclic recursive designs involving feedback loops are another possibility, such as $x \rightarrow m \rightarrow y \rightarrow x$, where $\hat{m} = f(x)$, $\hat{y} = f(m)$, and $\hat{x} = f(y)$. The last term represents a feedback loop, with a specified time interval, from $y$ to $x$. Empirical tests of cyclical recursive designs require a time-series analysis in which each variable is measured at a distinct time period that reflects the (causal) interval required for cause-effect relations to stabilize (cf. Heise, 1975; James et al., 1982; Strotz & Wold, 1971).

Finally, prior discussion has focused on complete mediation, where the antecedent $x$ affects the consequence $y$ only indirectly through the mediator $m$. The possibility of partial mediation also exists. A partial mediation model is usually displayed in one of the following two equivalent forms, given that relations are recursive:

$$x \rightarrow m \rightarrow y$$

$$x \rightarrow m \downarrow y$$

In these models, $x$ has both a direct effect on $y$ and an indirect effect on $y$, the latter being transmitted by $m$. This indicates that only part of the total effect of $x$ on $y$ is due to mediation by $m$ (cf. Duncan, 1970, 1975; Heise, 1975; Kenny, 1979). The mediation function has the form $\hat{m} = f(x)$, $\hat{y} = f(x, m)$. Analytic procedures for partial mediation models are over-viewed in Alwin and Hauser (1975).

**Summary**

There are many types of causal mediation relations and models. Yet, all have the common attribute that the mediator transmits influence from an antecedent to a consequence. The transmission need not involve all of the influence of the antecedent on the consequence, nor need the mediation relation be additive,
linear, or recursive. Indeed, many possible combinations exist. Nevertheless, each combination specifies a particular operational role for each variable, and mediators are those variables whose operational role involves transmission of influence. With mediators thus described, let us now turn to the question of specification errors in causal mediation models and tests of causal mediation models.

Objective 2: Identifying Specification Errors in Causal Mediation Models

A confirmatory test of a causal mediation model is designed to ascertain whether the model is useful for explaining how variables included explicitly in the model occurred and are related (cf. James et al., 1982). Confirmatory tests should only be conducted on “well-specified” causal models, by which it is meant that the assumptions or conditions for confirmatory analysis have been reasonably satisfied. Specification error is the general term used in confirmatory analysis to indicate that one or more conditions for confirmatory analysis has (have) not been reasonably satisfied. In the presentation below, we have selectively focused attention on specification errors considered to be of major salience in confirmatory tests of complete mediation models. The discussion is presented in the form of summary statements, with accompanying references because space considerations preclude furnishing a thorough review here. Brief mention is made of additional concerns in concluding remarks, with emphasis placed on the need to adopt analytic methods specifically designed for confirmatory analysis.

To preface our remarks, allow us to reiterate several points made in the introductory comments to this article. The specification errors discussed below are symptomatic of an incomplete transition from exploratory analysis to confirmatory analysis in areas such as job perception and attrition research. The specification errors became apparent only after knowledge accumulated concerning all of the conditions that are prerequisite to meaningful confirmatory analysis. In a sense, therefore, it is unfair to criticize prior research on complete mediation models inasmuch as researchers employed what at the time was considered a valid paradigm for causal analysis. On the other hand, a complete transition from exploratory analysis to confirmatory analysis will not be effected until the specification errors are recognized and subsequently addressed in future research. Thus, we will point out the specification errors in prior research, but we shall do so at a general level and in the interest of identifying the principles involved rather than raising ad hominem arguments in regard to specific studies.

Examples of important specification errors in causal, or structural, models based on complete mediation relations of the form \( x \rightarrow m \rightarrow y \) include the following: (a) misspecification of causal order (e.g., the true model is \( m \rightarrow x \rightarrow y \) or \( x \rightarrow y \rightarrow m \)); (b) misspecification of causal direction (e.g., the true model is \( x \rightarrow m \leftarrow y \)); (c) lack of self-containment, or an unmeasured variables problem, which is illustrated below; (d) the assumed additive, linear relations are nonadditive, nonlinear, or both; and (e) the model is unstable (i.e., nonstationary), which denotes that the variables and relations in the model are subject to severe random fluctuations or shocks (cf. James et al., 1982). It is only after (a) the model can be regarded as not being subject to one or more of these major specification errors and (b) after the model is shown by confirmatory analysis to have a good empirical fit with data that (c) it is justified to consider the results of the confirmatory analysis as useful for attempting to explain how a mediation process occurred (i.e., to make causal inferences), or to employ a term such as causal effect or its various euphemisms, such as “determine,” “indirect effect,” “influence,” and “transmit.”

Now consider that many mediation studies in the industrial and organizational literature and the organizational behavior literature begin with verbal, and often graphic, causal models in which considerable attention is given to causal order and explication of mediation processes. Causal relations are typically recursive throughout the model, although this appears to be more a matter of convenience than a well-thought-out, defensible case for unidirectional causation. Attention may or may not be given to additivity, linearity, and stability. However, attention is almost never given to the possibility of misspecification due to a “serious” unmeasured variables problem. By a serious unmeasured variables problem
is meant that a stable variable exists that (a) has a unique, nonminor, direct influence on an effect (either \( m \) or \( y \), or both); (b) is related at least moderately to a measured cause of the effect (e.g., is related to \( x \) in the functional equation for \( m \)); and (c) is unmeasured—that is, is not included explicitly in the causal model and the confirmatory analysis (James, 1980; James et al., 1982). This is unfortunate because a serious unmeasured variables problem precludes confirmatory analysis and the use of causal inference to attempt to explain mediation processes (cf. Billings & Wroten, 1978; Darlington, 1968; Duncan, 1970, 1975; James et al., 1982; Linn & Werts, 1969; Simon, 1952, 1953, 1977). In particular, confirmatory analytic techniques such as path analysis and structural equation analysis should not be used. If they are used, then, as shown in many of the references above, estimates of causal parameters and the ensuing causal inferences will be biased.

It is also the case that procedures typically associated with exploratory forms of analysis, namely hierarchical OLS or partial correlation, should not be employed in a confirmatory mode to test causal hypotheses or to serve as a basis for causal inference in the presence of a serious unmeasured variables problem. On the other hand, a serious unmeasured variables problem does not preclude the use of hierarchical OLS or partial correlation in an exploratory mode as long as the results of the hierarchical or partial correlation analysis are interpreted in correlational terms with no causal overtones. For example, an empirical test of a model of the form \( m = f(x) \), \( y = f(m) \), where the \( fs \) represent covariation and not causal relations, may be based on a hierarchical OLS and may show that \( R^2_{y,m} \) is not significantly greater than \( R^2_{y} \). This indicates that inclusion of \( x \) adds nothing to the prediction of \( y \) over that already furnished by \( m \). It may also be shown that \( R^2_{y,m} \) is significantly greater than \( R^2_{y,m} \), which denotes that \( m \) adds uniquely to the prediction of \( y \) in relation to \( x \). Such results support a correlational form of mediation and an interpretation such as the covariation between \( x \) and \( y \) vanishes if \( m \) is controlled." The results cannot be interpreted causally, such as \( m \) transmits causal influence from \( x \) to \( y \) or serves to explain how \( x \) and \( y \) are related, unless it can be assumed that the mediation relations are not subject to a serious unmeasured variables problem. Of course, use of correlational forms of mediation defeats the main purpose for developing and testing mediation models (i.e., explanation) and is the reason that most mediation models are presented in the causal mode (Rozeboom, 1956).

Unfortunately, it is often the case in field studies that causal mediation models with obvious misspecifications (i.e., unmeasured variables, unanalyzed reciprocal causation) have been subjected to goodness-of-fit tests using hierarchical OLS and/or partial correlation. In the context of present knowledge, these tests should be regarded as exploratory tests of correlational mediation hypotheses. Instead, these tests have been interpreted as confirmatory tests of causal hypotheses and used to make causal inferences. In effect, we have an unwarranted intertwining of confirmatory and exploratory procedures, which is evidenced by such things as the use of beta weights from the hierarchical OLS analyses as implicit path coefficients (i.e., the weights are interpreted in terms of importance and utility—cf. Darlington, 1968), the use of partial correlations that tend to zero (by controlling on a mediator) as evidence that an antecedent had no "direct effect" on a consequence, and the use of a significant increment in \( R^2_{y,x,m} \) in relation to \( R^2_{y,x} \) to support a causal inference that the "influence" of \( x \) on \( y \) is transmitted through the mediator \( m \). Remember also the example presented in the introduction to this article, where unmeasured variables would have to be invoked to attempt to explain within-job variation in job perceptions and job satisfaction.

In sum, the models, analyses, and results of many tests of mediation in the industrial and organizational literature and the organizational behavior literature do not furnish sufficient evidence for the causal interpretations offered in Discussion sections. It is recommended, therefore, that investigators begin to devote attention to all of the conditions for confirmatory analysis before conducting confirmatory tests on causal models and using the results of these tests to support causal inferences. A review of the conditions for confirmatory analysis and causal inference is presented in James et al. (1982). If one or more major sources of specification error is consid-
ered viable, then use procedures such as hierarchical OLS or partial correlation in the exploratory mode and limit discussion to correlational interpretations. On the other hand, if all sources of misspecification are considered, and no major misspecification is considered likely, then confirmatory analytic techniques such as path analysis and structural equation analysis should be used. This is because such techniques furnish (a) a means to test causal hypotheses that cannot be addressed by correlational techniques, such as reciprocal causation; (b) estimates of causal parameters; and (c) a basis for estimating "indirect effects," a major concern in mediation analysis (cf. Grif- fin, 1977). To illustrate the last point, if a model of the form \( x \rightarrow m \rightarrow y \) is confirmed, then the path coefficient linking \( x \) to \( m \) \( (p_{mx}) \) may be multiplied by the path coefficient linking \( m \) to \( y \) \( (p_{my}) \) or \( p_{mx}p_{my} \). This product reflects the magnitude of the indirect effect of \( x \) on \( y \). There is no analogue of this procedure in hierarchical OLS or partial correlation. On the other hand, we are not suggesting that hierarchical OLS and partial correlation have no place in confirmatory analysis. These methods have limited applications in the confirmatory mode (Cohen & Cohen, 1983), an example being the prior use of hierarchical OLS to test a portion of the causal hypotheses associated with a nonadditive, complete mediation model (see Figure 1). The point we wish to emphasize is that hierarchical OLS and partial correlation should not be used in place of confirmatory analytic methods.

Concluding Remarks

The following points were developed. First, mediation is generally thought of in terms of causal mediation, which connotes transmission of influences from antecedents to consequences and an attempt to explain how antecedents produce consequences. Second, mediation relations may assume any number of functional forms, including nonadditive, nonlinear, and nonrecursive forms. Third, confirmatory analytic techniques furnish the most informative tests of mediation. Fourth, there is no middle ground between exploratory (correlational) and confirmatory analysis. Attempts to explain how mediation processes occur by causal inference require well-specified causal models and empirically demonstrated goodness of fit between models and data. Specification errors, such as a serious unmeasured variables problem or misspecified causal direction, preclude confirmatory analysis and causal inference. Fifth, and finally, if causal models are well-specified, then confirmatory analytic techniques applicable to the model(s) of concern should be employed to furnish all relevant information (e.g., estimates of causal parameters, estimates of indirect effects, tests of nonadditivity, etc.).

A full treatment of mediation requires consideration of issues not addressed here. These issues include (a) the use of the intervening variable form of mediator in experimental analysis (MacCorquodale & Meehl, 1948; Rozeboom, 1956) and exploratory factor analysis (Royce, 1963); (b) the use of "mediating mechanisms" to develop theoretical rationales for causal hypotheses, whereby mediating mechanism means a hypothetical mediator that is not tested empirically (James et al., 1982); and (c) micromediation processes, which consist of mediating relations at a finer level of explanation than that of the model in question (e.g., at the level of receptor, neural, or muscular mediating processes—Cook & Campbell, 1979). Although important, these issues require a somewhat more esoteric presentation than the "applied" orientation of this article.


Received June 17, 1983
Revision received November 21, 1983.